

*This Book*  
*is*  
*Dedicated to the Memory*  
*of*  
SIR WILLIAM BATE HARDY  
(1864 — 1934)

JACQUES LOEB  
(1859 — 1924)

THOMAS BURR OSBORNE  
(1859 — 1929)

SØREN PETER LAURITZ SØRENSEN  
(1868 — 1939)

*Their investigations laid the foundations for the*  
*physical chemistry of proteins*

# PROTEINS, AMINO ACIDS AND PEPTIDES AS IONS AND DIPOLAR IONS

EDWIN J. COHN

AND

JOHN T. EDSALL

*Harvard Medical School*

INCLUDING CHAPTERS BY

JOHN G. KIRKWOOD

*Cornell University*

HANS MUELLER

*Massachusetts Institute of Technology*

J. L. ONCLEY

*Harvard Medical School*

*Massachusetts Institute of Technology*

GEORGE SCATCHARD

*Massachusetts Institute of Technology*



American Chemical Society  
Monograph Series

REINHOLD PUBLISHING CORPORATION

430 West Forty-second St., New York, U. S. A.



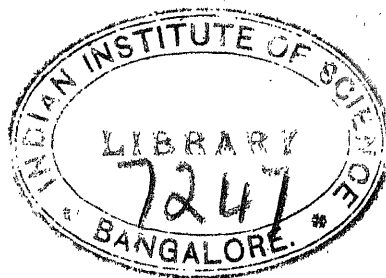
540.0202

N21.90

Copyright, 1948, by  
REINHOLD PUBLISHING CORPORATION

*All rights reserved.*

Second Printing





## GENERAL INTRODUCTION

### American Chemical Society Series of Scientific and Technologic Monographs

By arrangement with the Interallied Conference of Pure and Applied Chemistry, which met in London and Brussels in July, 1919, the American Chemical Society was to undertake the production and publication of Scientific and Technologic monographs on chemical subjects. At the same time it was agreed that the National Research Council, in coöperation with the American Chemical Society and the American Physical Society, should undertake the production and publication of Critical Tables of Chemical and Physical Constants. The American Chemical Society and the National Research Council mutually agreed to care for these two fields of chemical development. The American Chemical Society named as Trustees, to make the necessary arrangements for the publication of the monographs, Charles L. Parsons, secretary of the society, Washington, D. C.; the late John E. Teeple, then treasurer of the society, New York; and Professor Gellert Alleman of Swarthmore College. The Trustees arranged for the publication of the A. C. S. series of (a) Scientific and (b) Technologic Monographs by the Chemical Catalog Company, Inc. (Reinhold Publishing Corporation, successors) of New York.

The Council, acting through the Committee on National Policy of the American Chemical Society, appointed editors (the present list of whom appears at the close of this introduction) to have charge of securing authors, and of considering critically the manuscripts submitted. The editors endeavor to select topics of current interest, and authors recognized as authorities in their respective fields.

The development of knowledge in all branches of science, especially in chemistry, has been so rapid during the last fifty years, and the fields covered by this development so varied that it is difficult for any individual to keep in touch with progress in branches of science outside his own speciality. In spite of the facilities for the examination of the literature given by Chemical Abstracts and by such compendia as Beilstein's Handbuch der Organischen Chemie, Richter's Lexikon, Ostwald's Lehrbuch der

the knowledge on a given topic. Consequently when men who have spent years in the study of important subjects are willing to coördinate their knowledge and present it in concise, readable form, they perform a service of the highest value. It was with a clear recognition of the usefulness of such work that the American Chemical Society undertook to sponsor the publication of the two series of monographs.

Two distinct purposes are served by these monographs: the first, whose fulfillment probably renders to chemists in general the most important service, is to present the knowledge available upon the chosen topic in a form intelligible to those whose activities may be along a wholly different line. Many chemists fail to realize how closely their investigations may be connected with other work which on the surface appears far afield from their own. These monographs enable such men to form closer contact with work in other lines of research. The second purpose is to promote research in the branch of science covered by the monographs, by furnishing a well-digested survey of the progress already made, and by pointing out directions in which investigation needs to be extended. To facilitate the attainment of this purpose, extended references to the literature enable anyone interested to follow up the subject in more detail. If the literature is so voluminous that a complete bibliography is impracticable, a critical selection is made of those papers which are most important.

#### AMERICAN CHEMICAL SOCIETY

##### BOARD OF EDITORS

HARRISON E. HOWE, *Editor of Monographs*

##### Scientific Series:—

S. C. LIND,  
W. MANSFIELD CLARK,  
LINUS C. PAULING,  
L. F. FIESER.

##### Technologic Series:—

WALTER A. SCHMIDT,  
E. R. WEIDLEIN,  
F. W. WILLARD,  
W. G. WHITMAN,  
THOMAS H. CHILTON,  
BRUCE K. BROWN,  
W. T. READ,  
CHARLES ALLEN THOMAS.

## Preface

In this book we shall attempt a characterization of amino acids, peptides and proteins. It is our aim to examine the evidence concerning the size and shape of these molecules, and the number and distribution of the electric charges which they bear. Above all, we shall consider the implications of their charged structure for their physical properties, and their physico-chemical interaction with other molecules. We have not written a systematic treatise on all aspects of the chemistry of the proteins, or even of their physical chemistry, but have aimed only at presenting those subjects which are our primary concern from a definite, if limited, point of view. We have made no attempt to discuss experimental techniques, except when it has been necessary to do so in order to make clear the operational basis of the principles which are involved.

By the beginning of the present century, the fundamental work of many organic chemists had revealed a group of  $\alpha$ -amino acids as the essential constituents of proteins. Analytical evidence indicated that the proteins were large molecules, although their exact size remained undefined. During the past twenty-five years a revolution has occurred in the study of protein chemistry. By the use of osmotic pressure measurements and of the ultracentrifuge, it has been possible to determine the size of these very large molecules. X-ray diffraction studies have begun to reveal complex and systematic patterns in protein crystals, although the information at present obtainable from them is limited. The advance of organic chemistry has proceeded steadily. New and more powerful analytical techniques have been devised for determining the constituents of protein molecules. At least two essential amino acids of proteins have been discovered since 1920. New methods for the synthesis of peptides, and the study of their splitting by proteolytic enzymes, have deepened our knowledge of the chemical linkages in the protein molecule.

All these advances have influenced our thinking profoundly, and are discussed in the present monograph. However, we may point to two great papers which are landmarks in the approach to protein chemistry which is adopted here. Both of these papers were published in the same year, 1923. One is the treatment by Debye and Hückel of the interionic forces

of the work of Debye and Hückel, and of Bjerrum, are too far-reaching to permit a summary here. The greater part of this monograph represents an exploration of the field opened to investigation by these fundamental studies.

It was immediately apparent, when these papers appeared, that the study of dipolar ions and of interionic forces must be of great importance for the understanding of protein molecules. The complexity of the structure of proteins, however, raises great difficulties in the interpretation of experimental measurements made upon them. In consequence, it was desirable to turn to the simpler dipolar ions, the amino acids and peptides, of which the proteins are built, as models for the interpretation of the behavior of proteins. The first part of this monograph is entirely concerned with these simpler molecules, and forms an essential preliminary to the second part, which deals with the proteins. The general objectives of the whole book are more fully set forth in the first chapter.

The preparation of this book was first contemplated some fifteen years ago, when the senior author was invited by the Publication Committee of the American Chemical Society to prepare a monograph on the physical chemistry of the proteins. The foundations of the field of study concerned were, however, clearly in a state of flux at that time. New studies, first on amino acids and peptides and later on proteins, were undertaken as a necessary preliminary to any adequate treatment of dipolar ions. When we consider the present state of knowledge, we must recognize a profound ignorance of the details of the structural pattern of the protein molecule. Something approaching an adequate theory of the physical chemical behavior of amino acids and peptides, however, appears to have been achieved by the coöperative efforts of workers in several different laboratories. These studies on the simpler molecules have laid a foundation for the study of some aspects of the chemistry of proteins. Today, largely because of the development of the ultracentrifuge by Svedberg, we know much about the size of protein molecules and something about their shapes. Much is known about the number of the electrically charged groups carried by these molecules, and something about the distribution of these charges. Our present knowledge is still very far from the ultimate goal of protein chemists—that is, the complete elucidation of the structure of the protein molecule. The prospects of future advance, however, appear clearly in many directions and the possibilities of further development are unlimited.

This book has been written largely during a period of growing national emergency and was completed just at the time of the declaration of war.

found opportunity to include. We have omitted practically all discussion of protein surface films\* and of denaturation. Although the theory of electrophoresis is treated by Professor Mueller, the intended treatment of the experimental data in this field has been postponed.† Another field of importance, which must in the end influence our understanding of enzyme action, is the effect of dipolar ions on reaction velocities.‡ It is our hope to discuss all these topics at some future time.

This book represents the collaboration of a large group of closely associated workers including George Scatchard, John G. Kirkwood, Hans Mueller, J. L. Oncley, Ronald M. Ferry, W. T. Salter, Jeffries Wyman, Jr., A. von Muralt, Marcel Florkin, T. L. McMeekin, A. A. Green, J. P. Greenstein, N. F. Burk, N. R. Joseph, Danella Straup-Cope, V. E. Morgan, J. Steinhardt, J. A. Luetscher, Jr., H. L. Fevold, G. E. Perlmann, I. Fankuchen, W. L. Hughes, Jr., and L. E. Strong. Through constant association over a period of years, through discussions in seminar meetings, through coöperation in the laboratory, and through personal contacts, this body of thought has become the product of their collaboration. They have critically considered the experiments of previous investigators, and also the interpretation of new experiments in terms of older theories. When these seemed inadequate, new theoretical approaches have been attempted. Such discussions led, for instance, to the theory of interactions between ions and dipolar ions developed by Scatchard and Kirkwood, and later extended by Kirkwood.

The scope of the field is such that one or two authors could not do justice to it. We are fortunate in that George Scatchard§, John G. Kirkwood, Hans Mueller and J. L. Oncley¶ have prepared the fundamental chapters concerning subjects to which they have made essential contributions. The collaboration of J. W. Mehl in Chapter 18 and of J. D. Ferry in Chapter 24 has also been of great value, as have been the revisions to Chapter 15, suggested by H. B. Vickery. Several of our other colleagues contributed to this book tables or data concerning subjects with which they are particularly familiar; all have contributed invaluable aid in the developments recorded.

We are also greatly indebted to Mrs. Anna B. Curtiss for her steady perseverance, over a period of years, in preparing this manuscript for publication. Her careful and arduous work in checking the tables and bibliog-

\* See I. Langmuir, *Cold Spring Harbor Symposium on Quantitative Biology*, 6, 171 (1938).

† The literature of this field up to 1934 is covered in the monograph by H. A. Abramson, "Electrokinetic Phenomena", New York, Chem. Cat. Co., (Reinhold Publishing Corp.) 1934. More recent developments are well surveyed in a symposium published in the *Annals of the New York Academy of Sciences*, Volume 39, pp. 105-212 inclusive, 1939. See also Abramson, H. A., *Mem. Am. Chem. Soc.*, 1, 1, and Gorin, M. H., "Electrophoresis of Proteins", Reinhold Publishing Corp., New York, 1941. See also North, L. G., *Chem. Rev.*, 30, 323 (1942).

‡ See, for instance, Straup, D., and Cohn, E. J., *J. Am. Chem. Soc.*, 57, 1794 (1935); Scatchard, G., *J. Chem.*

raphy, the text and figures and in reading the proof has been an indispensable aid.

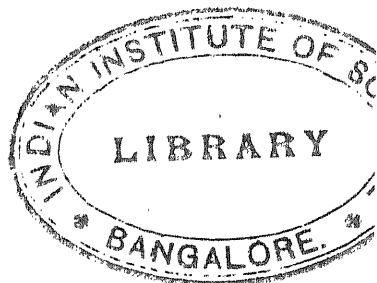
The junior author is greatly indebted to the John Simon Guggenheim Memorial Foundation for a fellowship which permitted him to work at the California Institute of Technology during 1940-41. During this period he wrote or revised large portions of the present work and profited greatly from discussions there with members of the staff, particularly with Linus Pauling, Carl Niemann and Robert B. Corey.

The continued support of the Rockefeller Foundation over a period of years, given without defining the project in advance, has permitted the natural development of these investigations. Freed from the necessity of considering immediate practical applications, it was possible for a group of investigators to make the excursion into the chemistry of protein prototypes which forms the background of this book and which, in turn, we hope may yield a deeper understanding of the proteins.

EDWIN J. COHN  
JOHN T. EDSALL

*Harvard Medical School  
Boston, Massachusetts  
December, 1941*





## Contents

GENERAL INTRODUCTION.....	v
---------------------------	---

<i>Chapter</i>	<i>Page</i>
1. INTRODUCTION.....Edwin J. Cohn and John T. Edsall	1

### Part I

#### Amino Acids and Peptides

2. SPECTROSCOPY AND DIPOLAR IONIC STRUCTURE..John T. Edsall	9
Vibrational Spectra.....	9
Raman Spectra.....	10
The Ionization of the Carboxyl Group.....	12
The Ionization of the Amino Group.....	14
Infrared Spectra.....	19
Ultraviolet Spectra.....	19

3. THERMODYNAMICS AND SIMPLE ELECTROSTATIC THEORY.	
George Scatchard	20
Introduction.....	20
The Chemical Potential.....	22
The Perfect Gas.....	23
Real Gases.....	23
Dilute Solutions.....	25
Ideal Solutions.....	27
Chemical Equilibria.....	29
The Measurement of Chemical Potentials.....	31
Unrestricted Systems.....	36
Liquid Junction Potentials.....	39
Single Ion Activities, Salt Bridges and pH.....	42
Osmotic Pressure.....	45
Donnan Equilibrium.....	46
Membrane Potentials.....	47
Gravitational and Centrifugal Fields.....	48
Surface Tension.....	49
Solubility and Particle Size.....	51
Simple Electrostatic Theory.....	52
The Debye-Hückel Equation.....	54
Aqueous Solutions.....	59
Heat and Volume Changes.....	60

<i>Chapter</i>		<i>Page</i>
	Güntelberg-Müller Charging Process.....	66
	Chemical Equilibria.....	67
	Chemical Kinetics.....	72
4.	DIPOLAR IONS AND ACID-BASE EQUILIBRIA John T. Edsall	75
	Thermodynamic Dissociation Constants of the Amino Acids.....	78
	Dissociation Constants from Cells with Liquid Junction.....	83
	Apparent Heats of Ionization of Amino Acids and Peptides.....	88
	Isoelectric Points of Amino Acids and Peptides.....	90
	Net Charge and Total Charge on Ampholyte Molecules: The Point of Maximum Charge.....	93
	The Equilibrium between Dipolar Ions and Uncharged Molecules....	96
	The Effect of Temperature on $K_z$ .....	104
	The Relative Concentrations of Dipolar Ions and Uncharged Molecules in Solvents of Lower Dielectric Constant.....	105
	Relative Acidity of Acids of Different Charge Types in Different Media	105
	The Effect of Varying Ionic Strength on Acid-Base Equilibria Involving Dipolar Ions.....	111
5.	SOME RELATIONS BETWEEN ACIDITY AND CHEMICAL STRUCTURE John T. Edsall	116
	Effect of Substituent Groups on the Acidity of Neighboring Groups	116
	Effect of Charged Groups.....	117
	Dipolar Substituents.....	120
	The Resonance Effect.....	124
	Resonance in the Amide Group.....	129
	The Strength of Certain Other Acid Groups.....	130
	The Aromatic Hydroxyl (Phenolic) Group.....	130
	The Sulfhydryl Group, Aliphatic and Aromatic.....	131
	The Imidazole and Pyrazole Groups.....	131
	The Guanidonium (Guanidinium) Group.....	132
	The Phosphoric Ester Group.....	133
	Quaternary Ammonium Groups in Dipolar Ions.....	134
	The Effect of the Peptide Linkage on Acidic Dissociation.....	134
	The Acidity of Some Ring Structures.....	135
	The Effect of Formaldehyde upon the Titration Constants of Amino Acids.....	136
6.	DIELECTRIC CONSTANTS AND DIPOLE MOMENTS OF DIPOLAR IONS..... John T. Edsall	140
	Definition and Measurement of the Dielectric Constant.....	140
	Relation of Dielectric Constant to Molecular Structure:	
	Polar and Non-polar Molecules.....	141
	The Distortion Effect.....	142
	The Orientation Effect.....	142

# CONTENTS

xiii

Chapter	Page
7. APPARENT MOLAL VOLUME, HEAT CAPACITY, COMPRESSIBILITY AND SURFACE TENSION OF DIPOLAR IONS IN SOLUTION	John T. Edsall 155
Apparent Molal Volumes.....	157
Electrostriction Due to Charged Groups in Ions and Dipolar Ions..	157
Apparent Molal Volume as a Function of Concentration.....	161
Apparent Molal Heat Capacities.....	165
Apparent Molal Compressibility.....	173
Surface Tensions of Amino Acid Solutions.....	174
8. SOLUBILITY OF AMINO ACIDS, PEPTIDES AND RELATED SUBSTANCES IN WATER AND ORGANIC SOLVENTS	John T. Edsall and George Scatchard 177
Solubility and Intermolecular Forces.....	177
Polar Liquids and Hydrogen Bonds.....	180
Melting Points and Crystal Structure of Amino Acids; Their Signifi- cance for Solubility.....	184
Solubility and Heat of Solution of Amino Acids in Water.....	187
9. INTERACTIONS BETWEEN ORGANIC SOLVENTS AND DIPOLAR IONS ESTIMATED FROM SOLUBILITY RATIOS	Edwin J. Cohn and John T. Edsall 196
Relative Solubility in Water and in Organic Solvents.....	196
Effect of the CH <sub>2</sub> Groups of Solutes on Solubility Ratios and on Surface Tensions.....	205
Effect of the CH <sub>2</sub> Groups of Solvents on Solubility Ratios and Activity Coefficients.....	210
The Effect of Various Other Groups on Solubility Ratios and Activity Coefficients.....	211
Influence of Dipolar Ionization on Solubility Ratios and Activity Coefficients.....	212
Solubility in Ethanol-Water Mixtures.....	214
10. INTERACTIONS BETWEEN AMINO ACIDS, PEPTIDES AND RELATED SUBSTANCES.....	Edwin J. Cohn 217
Activity Coefficients of $\alpha$ -Amino Acids as a Function of Concentration	217
Influence of CH <sub>2</sub> Groups.....	219
Influence of OH Groups.....	221
Activity Coefficients of Asparagine as a Function of Concentration of Other Dipolar Ions.....	222
Influence of CONH Groups.....	222
Influence of CH <sub>2</sub> Groups.....	225
Activity Coefficients of Cystine as a Function of Concentration of	

<i>Chapter</i>	<i>Page</i>
Activity Coefficients of Peptides as a Function of Concentration.....	226
Interaction between Dipolar Ions.....	228
Theoretical Models.....	231
<b>11. INTERACTIONS BETWEEN IONS AND AMINO ACIDS OR PEPTIDES</b>	
Edwin J. Cohn	236
The Solubility of Dipolar Ions in Aqueous Salt Solutions.....	237
Solubility of Asparagine in Aqueous Salt Solutions.....	239
Solubility Ratios and Activity Coefficients of Cystine in Aqueous Salt Solutions.....	241
Activity Coefficients of Cystine in Aqueous Salt Solutions Containing Other Amino Acids.....	246
Solubility of Salts in Aqueous Solutions Containing Dipolar Ions....	251
Influence of the Salt.....	251
Influence of the Dipolar Ions.....	253
Freezing Point and Electromotive Force Measurements in Aqueous Salt Solutions Containing Dipolar Ions.....	256
Decrease in Activity Coefficient of the Salt.....	257
Increase in Activity Coefficient of the Salt.....	259
Increase in Activity Coefficient of Dipolar Ions.....	260
Decrease in Activity Coefficient of Dipolar Ions.....	261
Sodium Chloride and Glycine.....	261
Potassium Chloride and Glycine.....	264
Zinc Chloride and Amino Acids.....	264
Interactions between Ions and Dipolar Ions in Alcohol-Water Mixtures	266
Glycine and Sodium Chloride.....	266
Glycine and Other Alkali Halides.....	269
Glycine and Lithium Chloride.....	270
$\alpha$ -Aminocaproic Acid and Lithium Chloride.....	271
$\epsilon$ -Aminocaproic Acid and Lithium Chloride.....	272
Hydantoic Acids, Peptides and Sodium Chloride.....	273
<b>12. THE THEORETICAL INTERPRETATION OF THE PROPERTIES OF SOLUTIONS OF DIPOLAR IONS.....</b>	<b>John G. Kirkwood 276</b>
Thermodynamic Interaction with Electrolytes.....	277
Acid-Base Equilibrium.....	287
Dielectric Polarization.....	294
Appendix to Section 2.....	298
<b>Part II</b>	
<b>Proteins</b>	
<b>13. THE STRUCTURAL BASIS OF THE PROTEIN MOLECULE: EVIDENCE FROM ANALYSIS AND FROM THE ACTION OF PROTEOLYTIC ENZYMES.....</b>	<b>John T. Edsall 307</b>

# CONTENTS

xv

<i>Chapter</i>	<i>Page</i>
The Nature of the Peptide Linkages and the Free Acid and Basic Groups of Proteins.....	311
The Protamines, and the Structure of Clupein.....	312
Cosubstrates in Proteolysis.....	314
The Hypothesis of Bergmann and Niemann.....	315
14. X-RAY DIFFRACTION STUDIES AND PROTEIN STRUCTURE	
John T. Edsall	318
Interatomic Distances in Amino Acids and Diketopiperazine.....	318
Inferences Concerning the Structure of Polypeptide Chains.....	321
The Fibrous Proteins.....	322
The "Globular" Proteins.....	327
Virus Proteins.....	332
Tentative Inferences Concerning Protein Structure.....	333
The Cyclol Hypothesis.....	334
Arrangement of Protein Molecules in Crystals: Patterson Projections	336
15. THE ELEMENTARY AND AMINO ACID COMPOSITION OF THE PROTEINS.....	Edwin J. Cohn 338
Discovery of Amino Acids.....	338
Sulfur-Containing Amino Acids.....	340
Phosphorus and the Hydroxyamino Acids.....	340
Rare Amino Acids.....	342
Classes of Amino Acids.....	344
Analysis of Amino Acids.....	346
Sulfur-Containing Amino Acids.....	346
Basic Amino Acids.....	352
Tryptophane and Tyrosine.....	356
Aliphatic Amino Acids.....	361
Amino Acid Composition of Proteins.....	366
16. DENSITY AND APPARENT SPECIFIC VOLUME OF PROTEINS	
Edwin J. Cohn and John T. Edsall	370
The Specific Volumes of Amino Acid Residues.....	371
The Glycyl Residue, $-\text{NH}\cdot\text{CH}_2\cdot\text{CO}-$ .....	371
Correction for Covolume and Electrostriction.....	373
Specific Volumes of Proteins.....	374
Partial Specific Volumes of Proteins in Solution.....	375
Density of Protein Crystals.....	377
17. OSMOTIC PRESSURE AND MOLECULAR WEIGHTS OF PROTEINS	
John T. Edsall and Edwin J. Cohn	382
18. TRANSLATIONAL DIFFUSION OF AMINO ACIDS AND PROTEINS	
John T. Edsall and John W. Mehl	396

<i>Chapter</i>	<i>Page</i>
Diffusion and Brownian Movement.....	402
Experimental Methods.....	406
The Diffusion Coefficients of the Amino Acids.....	410
The Diffusion Coefficients of the Proteins.....	411
 19. SEDIMENTATION AND DIFFUSION IN CENTRIFUGAL FIELDS:	
MOLECULAR WEIGHTS OF PROTEINS.....John T. Edsall	419
Effect of a Centrifugal Field in Modifying the Diffusion Process.....	419
Sedimentation Equilibrium.....	420
Sedimentation Velocity.....	421
Frictional Ratios and Their Relation to Asymmetry and Hydration..	424
Experimental Methods.....	426
Results of Ultracentrifugal Studies on Proteins.....	427
Molecular Weight Classes.....	431
Comparison with Osmotic Data.....	433
Frictional Ratios, Molecular Shape, and Hydration.....	434
Effects of Denaturing Agents.....	435
Dissociation and Association of Protein Molecules.....	437
 20. PROTEINS AS ACIDS AND BASES.....John T. Edsall	444
Net Charge, Total Charge and Maximum Acid and Base Binding....	446
Determination of Nature of Ionizing Groups in Proteins.....	447
Titration Constants of Polyvalent Acids and Bases.....	449
Calculation of Free and Bound Acid and Base.....	453
Equilibria between Different Ionic Forms in Polyvalent Acids and Ampholytes.....	460
The Equation of Linderstrøm-Lang.....	462
The Probability Distribution Function for a System of Independently Ionizing Groups.....	463
A Simple Case.....	464
Equilibria between Different Charged Forms in Isoelectric Hemoglobin	465
The Effect of Ionic Strength on Protein Titration Curves.....	468
The Titration Curves of Certain Important Proteins.....	477
Hemoglobin.....	478
Maximum Acid and Base Combining Capacity.....	478
The Titration Constants of Hemoglobin.....	479
Heat of Ionization of the Acid and Basic Groups in Hemoglobin..	481
The Heme-Linked Acid Groups of Hemoglobin.....	482
Cytochrome C.....	486
Zein.....	489
The Titration Curve of Gelatin: Influence of Formaldehyde and of Alcohol.....	492
Effects of Deamination.....	494
Wool Protein: Effect of Different Acid Anions on the Titration Curve	495
Egg Albumin.....	498

# CONTENTS

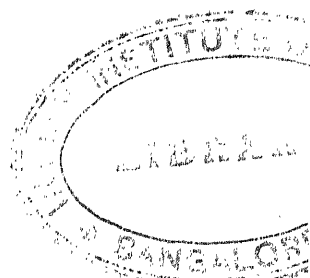
xvii

<i>Chapter</i>	<i>Page</i>
21. ROTARY BROWNIAN MOVEMENT. THE SHAPE OF PROTEIN MOLECULES AS DETERMINED FROM VISCOSITY AND DOUBLE REFRACTION OF FLOW.....John T. Edsall	506
Rotary Diffusion Constants and Relaxation Times.....	507
Relation to Molecular Size and Shape.....	510
The Viscosity of Protein Solutions.....	512
Measurement of Viscosity.....	513
The Viscosity of Solutions of Large Molecules.....	514
Rotary Motion of Ellipsoidal Particles Subjected to a Velocity Gradient.....	517
Relation of Viscosity Increment to Shape of Molecules.....	519
Electroviscous Effects.....	523
Change of Viscosity with Velocity Gradient: "Non-Newtonian" Flow.....	523
Double Refraction of Flow in Protein Solutions.....	527
Concentric Cylinder Apparatus: The Cross of Isocline and the Extinction Angle.....	528
Double Refraction of Ellipsoidal Molecules in Parallel Orientation.....	531
Theory of Double Refraction of Flow: Determination of Rotary Diffusion Constant.....	532
Effect of Chemical Agents on Double Refraction of Myosin.....	536
Tobacco Mosaic Virus.....	537
Influence of Polydispersity.....	539
Studies on Other Proteins: The Lengths of Protein Molecules.....	541
22. THE ELECTRIC MOMENTS AND THE RELAXATION TIMES OF PROTEINS AS MEASURED FROM THEIR INFLUENCE UPON THE DIELECTRIC CONSTANTS OF SOLUTIONS.....J. L. Oncley	543
Simple Theory.....	543
Calculation of Dipole Moments.....	546
Methods of Measurement.....	547
Electrode Polarization.....	549
Dielectric Increments.....	550
Low Frequency Dielectric Increments.....	550
High Frequency Dielectric Increment.....	553
Total Dielectric Increment.....	554
Dielectric Dispersion.....	555
Dispersion of Conductance.....	558
Results.....	560
Carboxyhemoglobin (horse).....	560
Carboxyhemoglobin (pig).....	561
Myoglobin.....	561
Insulin.....	561
Lactoglobulin.....	563
Egg Albumin.....	563

ERRATA

<i>Chapter</i>	<i>Page</i>
Secalin.....	565
Zein.....	565
Amino Acids and Peptides.....	565
Conclusion.....	567
23. THE SOLUBILITY OF PROTEINS..... Edwin J. Cohn	569
I. The Separation and Classification of Proteins.....	569
II. Differences in the Solubility Relations of Different Proteins.....	573
III. Application of the Phase Rule to the Solubility of Proteins	
John T. Edsall	576
The Thermodynamic Activity of the Protein in the Crystalline	
Phase.....	584
24. THE INTERACTIONS OF PROTEINS WITH IONS AND DIPOLAR IONS	
Edwin J. Cohn and John D. Ferry	586
I. Solubility and the Heat of Solution.....	586
II. Solubility and Equivalent Combining Weights.....	588
III. Solubility and the Heat of Neutralization.....	590
IV. Solubility Product Constant.....	593
V. Solubility and the Activity Coefficients of Protein Ions.....	594
VI. Solubility in Systems Containing Insoluble Protein Salts.....	599
VII. Solubility in Concentrated Salt Solutions: Salting-Out Values	
of $K_s$ .....	603
$\beta$ as a Function of pH.....	606
$\beta'$ as a Function of Temperature.....	607
VIII. Solubility in Dilute Salt Solutions: Salting-In.....	609
Hemoglobin.....	612
Lactoglobulin.....	614
Influence of Dipole Moment on Salting-In.....	615
Influence of Protein Ions on Salting-In.....	616
IX. Solubility in Solutions Containing More than One Dipolar Ion....	617
Hemoglobin.....	617
Lactoglobulin.....	619
X. Interactions in Protein Solutions Calculated from Electromotive	
Force and Osmotic Measurements.....	619
25. THE THEORY OF ELECTROPHORETIC MIGRATION... Hans Mueller	623
The Electrophoretic Flow Parallel to Plane Surfaces.....	626
The Electrophoretic Migration of Spherical Particles.....	631
APPENDIX. DIELECTRIC INCREMENTS OF ELLIPSOIDAL MOLECULES	
OF DIFFERENT AXIAL RATIOS.....	644
TABLE OF SYMBOLS.....	652
SUBJECT INDEX.....	659





## Chapter 1

### Introduction

BY EDWIN J. COHN AND JOHN T. EDSALL

Molecules are of several widely different types. At one extreme are very volatile, non-polar substances, such as the rare gases and the aliphatic hydrocarbons; at the other are the totally ionized salts such as sodium chloride, consisting of ions which migrate in an electric field, and exert intense electrostatic forces on their neighbors. Intermediate between these two extremes is the great class of polar molecules. A polar molecule does not give rise to ions capable of migration in a uniform electric field; but because of its polarity it does tend to orient itself in such a field. Polar molecules also exert more powerful attractive forces on their neighbors than do non-polar molecules; correspondingly, they are less volatile and boil at a higher temperature than non-polar molecules of similar size. Alcohols, amines, esters, ketones, nitriles, organic halogen derivatives—indeed the great majority of organic compounds, apart from the hydrocarbons—are polar molecules. So also are a large proportion of inorganic molecules.

Among the compounds which ionize in water there are two great classes—the strong electrolytes, which are completely ionized in water, and the weak electrolytes, which at moderate concentrations are only slightly ionized. Relatively few electrolytes are in the class which is intermediate between these two extremes. Trichloroacetic acid, phosphoric acid (considering only the first step in its dissociation), and the  $\text{HSO}_4^-$  ion are examples of such electrolytes of intermediate strength.

Polar molecules also fall into two great classes, one class having relatively low, the other extremely high, electric moments. The former includes those groups of substances already mentioned—the alcohols, amines, esters, ketones, nitriles and the like. These compounds are relatively volatile, and the intermolecular forces are not very intense, although far stronger than between non-polar molecules. The second class of polar molecules includes aliphatic amino acids, peptides, proteins, betaines, and certain phospholipids. Typical compounds of this class, for example, the am-

polar solvents, and are relatively far more soluble in water or in salt solutions. Many of them, for instance glycine, betaine,  $\beta$ -alanine,  $\epsilon$ -aminocaproic acid, and the albumins, dissolve in water to form extremely concentrated solutions. In short, even though they are electrically neutral, they display many properties which render them akin to the totally ionized salts. When dissolved in water, they give solutions of dielectric constant much higher than that of water itself; whereas almost all ordinary polar molecules, when dissolved in water, decrease the dielectric constant.

All these phenomena indicate that such molecules must be surrounded by very intense fields of force. Their kinship to true ions has led to their being termed in German *Zwitterionen*, and, in English, dipolar ions. We shall use the latter term to describe them throughout this book.<sup>1</sup>

The quantitative index of the polarity of a molecule is its electric moment (Chapters 6, 12, and 22). The order of magnitude of the moments to be expected in simple molecules may be readily estimated. Consider a proton and an electron, both carrying a charge of magnitude  $4.8 \times 10^{-10}$  electrostatic units, separated by a distance of one Ångström ( $10^{-8}$  cm.). The moment of such a system is the product of charge and distance. In this case it is  $4.8 \times 10^{-10} \times 10^{-8} = 4.8 \times 10^{-18}$  e.s.u., or 4.8 Debye units. Nearly all molecules which are not dipolar ions possess electric moments smaller than this. The electric moments of water and the alcohols are somewhat less than 2 Debye units, those of ketones near 2.7, those of nitriles and organic nitro compounds between 3 and 4. On the other hand, glycine,  $^+\text{H}_3\text{N} \cdot \text{CH}_2 \cdot \text{COO}^-$ , the simplest of the dipolar ions, has an electric moment of approximately 15 Debye units. This value may readily be estimated from the magnitude of the charge on the ammonium and carboxylate groups, and from the distance between them, which is known from x-ray diffraction studies (Chapters 8 and 14) to be approximately three Ångström units. This value has been confirmed, within a relatively small limit of probable error, by several independent types of measurement (See especially Chapters 6, 10, 11 and 12). Measurements of the same sort lead to much higher moments for such substances as  $\epsilon$ -aminocaproic acid, glycylglycine, and other peptides. The proteins are found to have dipole moments of many hundreds, or even thousands, of Debye units (Chapter 22). Thus the compounds here considered form a special group, with properties for which there are few analogs in other realms of chemistry.

All dipolar ions may be converted, by the addition of acids or bases, into

<sup>1</sup> The term "dipolar ion" was apparently first used by Ingold, C. K., *Chem. Rev.*, 15, 225 (1934). We have adopted it as perhaps the most satisfactory equivalent of the German *Zwitterion*. Several other terms such

ions carrying a net charge. Thus glycine in acid solutions is converted into the cation  $^+\text{H}_3\text{N}\cdot\text{CH}_2\cdot\text{COOH}$ , and in alkaline solution into the anion  $\text{H}_2\text{N}\cdot\text{CH}_2\cdot\text{COO}^-$ . Protein molecules, carrying great numbers of acid and basic groups, can form a great variety of ions of differing net charge. The study of acid-base equilibria, which determine the relative amounts of all these ions and dipolar ions under given conditions, is therefore fundamental to the understanding of the properties of molecules of this class (See especially Chapters 4, 5 and 20).

That isoelectric amino acids and peptides exist as dipolar ions in solution is revealed with particular clarity by studies of Raman and infrared spectra (Chapter 2). There is scarcely a property of these molecules, however, which is not conditioned by their dipolar ionic structure. Their solutions in water show higher densities, and smaller heat capacities, than those of most similar organic compounds. These effects are presumably due to electrostriction of the solvent, produced by the charged groups of the dipolar ions (Chapter 7). Owing to their great electric moments, they give solutions with the highest dielectric constants yet known (Chapter 6). For example, the dielectric constant of a saturated solution of  $\epsilon$ -aminocaproic acid is probably near 300, and other dipolar ions can be prepared which should give values well above this. The production of media with dielectric constants of this magnitude opens a new field of research in electrochemistry. The biochemical implications of these phenomena may be far-reaching, since all living cells contain proteins in concentrations sufficient to affect greatly the dielectric constant within the cell.

The thermodynamic properties of dipolar ions, owing to the charged groups which they contain, are profoundly affected by the electrical properties of the surrounding medium, in particular by the concentration of ions present, the valence of the ions, and the dielectric constant of the solvent. The thermodynamic relations involved, and the principles of the simpler electrical interactions, are treated by Scatchard (Chapter 3). The thermodynamic properties of model structures corresponding closely to actual dipolar ions are presented in detail by Kirkwood (Chapter 12). We have devoted several chapters to the consideration of solubility studies, and of other methods of determining activity coefficients of amino acids and peptides (Chapters 8, 9, 10, 11). The comparison of experiment and theory for simple dipolar ions in these chapters forms an essential preliminary to the consideration of similar studies on proteins (Chapters 23 and 24).

That some molecules must be dipolar ions was first inferred by Bredig<sup>2</sup> in the case of betaine, and by Küster<sup>3</sup> in the case of the acid (red) form of

methyl orange. For a long period, however, it was not realized that a structure of this type should be assigned to the aliphatic amino acids. The structure of isoelectric glycine was traditionally formulated as  $\text{H}_2\text{N}\cdot\text{CH}_2\cdot\text{COOH}$ , and the state of the amino and carboxyl groups in other isoelectric amino acids and peptides was expressed in the same manner. In 1915 E. Q. Adams<sup>4</sup> pointed out that the dissociation constants of the amino acids were inexplicable except on the assumption of dipolar ionic structure; but his views did not at the time receive the attention they deserved. In 1922 Pfeiffer<sup>5</sup> showed clearly, from steric considerations and melting points, that the betaines must be dipolar ions, and inferred that the amino acids must possess a similar structure. Independently of Pfeiffer's work, the masterly analysis of Bjerrum<sup>6</sup> brought to a focus all the evidence then available. Bjerrum pointed out that in the amino acids an equilibrium must exist in solution between a dipolar ion of the structure  $^+\text{H}_3\text{N}\cdot\text{CH}_2\cdot\text{COO}^-$  and its isomer  $\text{H}_2\text{N}\cdot\text{CH}_2\cdot\text{COOH}$ . He showed that in the aliphatic amino acids, the dipolar ion is the form almost exclusively present. The appearance of this classic paper marked a new epoch in the history of this field. Bjerrum's conclusions were confirmed and extended by workers in many other laboratories; measurements of dielectric constants and of Raman and infrared spectra furnished the most striking of all the proofs for the theory; and the dipolar ion concept has now been recognized as fundamental to the understanding of the behavior of amino acids, peptides and proteins.

The first part of this book is devoted to the simpler dipolar ions—amino acids, peptides, betaines and phospholipids; the more complicated questions arising in the study of the proteins are treated in the second part. In many respects, however, we have approached the study of proteins from the same standpoint adopted in dealing with the amino acids and peptides.

The structural pattern of protein molecules is not known in detail. Certain general features of structure can, however, be regarded as established (Chapters 13 and 14), and much is known about the relative amounts of different amino acids which proteins yield on hydrolysis (Chapter 15). But the greatest problems in the structural chemistry of proteins are still unsolved. Even concerning the size of protein molecules no accurate estimates could be made as recently as thirty years ago. Since that time the study of osmotic pressures and of diffusion coefficients, and the development of the ultracentrifuge, has placed our knowledge of the size of protein molecules upon a firm foundation (See Chapters 16, 17, 18, 19).

Proteins are highly specialized molecules; the general elements in the fundamental patterns of structure are capable of wide variation in order

to adapt them to their biological function, giving molecules very different in size, shape and chemical composition. Protein molecules may be nearly spherical, like egg albumin and hemoglobin; they may be very thin elongated structures, such as the fibroin of silk, the keratin of hair or wool, and the myosin of muscle. Knowledge of the shapes of protein molecules is today less exact than knowledge of their sizes, but evidence concerning shapes is now available from several independent methods of measurement.

For proteins in the solid state, x-ray diffraction studies (Chapter 14) provide the most important information concerning molecular shape. For proteins in solution, the available evidence is derived from sedimentation and diffusion measurements (Chapters 18 and 19), determinations of viscosity and double refraction of flow (Chapter 21), and dielectric constant measurements carried out over a wide range of frequencies (Chapter 22).

The very great solvent and precipitating action of salts and organic solvents on proteins, and the interaction of one protein with another, are largely determined by the charged groups carried by all protein molecules. Qualitatively the phenomena are similar to those noted in solubility studies on the amino acids and peptides. Quantitatively the effects found in protein systems are vastly greater. This is a natural consequence of the great size and large electric charge on protein molecules.

The migration of proteins in an electric field, or electrophoresis, has been of fundamental importance in their study ever since the work of Hardy more than forty years ago. The relation of the electric mobility to the charge carried by the molecule, however, is difficult to formulate, and many points in the theory have remained obscure. The underlying theory is critically examined by Professor Mueller in the concluding chapter (Chapter 25).

The structure and the functions of the proteins in the living cell are multi-form. Among proteins are included the carriers of oxygen to the tissues; the structural elements of muscle fiber, which are responsible for its contraction; the carriers of immune reactions in the blood; hormones, such as insulin and the pituitary hormones; bacteriophage; and many at least of the viruses of plant and probably of animal diseases. Such enzymes as pepsin, trypsin, chymotrypsin, carboxypeptidase, urease, catalase, and the protein components of certain dehydrogenase systems, have all been isolated as crystalline proteins; and the exploration of this field has only begun.

In addition to protein dipolar ions, phospholipids are essential components of all cells, and their function is particularly prominent in the brain and nervous system. These phospholipids differ in structure, but all

presence of large fatty acid radicals in these molecules, along with the charged groups which make them dipolar ions, gives them a strong attraction both for fatty solvents and for water. These properties must be intimately related to their function in the cell.

The study of the physical chemistry of hemoglobin has revealed with particular clarity and beauty the relation between the chemistry of a protein molecule and its specialized biological function. As other proteins become better known, we may confidently expect that similar and equally specific adaptations will be revealed. The investigation of this borderline between chemistry and biology must lead ultimately to the discovery of underlying relations equally fundamental for both sciences.

## Part I

# Amino Acids and Peptides





## Chapter 2

# Spectroscopy and Dipolar Ionic Structure

BY JOHN T. EDSALL

By the uptake of energy from their surroundings, molecules may be set into rotation or vibration, or may undergo electronic transitions. Spectroscopic data furnish the most direct evidence concerning the alterations of molecular energy involved in such changes of state, since the spectroscopic frequencies observed are directly proportional to the energy changes. Thus the ionization of a certain group within a molecule is always accompanied by changes in the molecular spectrum, which in general are highly characteristic of the group involved. If a neutral aminocarboxylic acid exists as a dipolar ion, both the amino and the carboxyl group are ionized. The charge on one group can be removed by adding acid, on the other by adding alkali. Thus by studying the spectra of the free amino acid and its acid and alkaline salts, and comparing them with the spectra of amines, carboxylic acids, and their salts, the state of the isoelectric amino acid may be deduced.

### Vibrational Spectra

The spectra giving the most useful data on this point are infrared and Raman spectra, arising from molecular vibrations. Such vibrations are very closely related to the structural pattern of the atoms in the molecule, and to the strength of the bonds between them<sup>1</sup>. We assume that the forces between atoms linked by a chemical bond obey Hooke's law—that is, when the atoms are displaced from their equilibrium distance  $r_0$ , by an amount  $r - r_0$ , the restoring force,  $f$ , is proportional to the displacement;  $f = F(r - r_0)$ , where  $F$  is the force constant of the bond (generally expressed in dyne/cm.). This is in general true only when  $r - r_0$  is small compared to  $r_0$ ; but in most cases that arise in practice it is a good approximation. Thus a diatomic molecule, made up of two atoms with masses  $m$  and  $M$ , vibrates with only one fundamental frequency  $\nu = (1/2\pi)\sqrt{F/\mu}$  where  $\mu$  (the "reduced mass") is given by the relation  $1/\mu = (1/m) + (1/M)$ . The greater the force constant,  $F$ , of the bond, the higher the frequency

triple bonds. Thus there is a fairly close proportionality between  $F$  and the character of the bond<sup>2</sup>.

A bent symmetrical triatomic molecule ( $Mm_2$ ), such as water, has three types of fundamental vibration (Fig. 1). The system is characterized by the masses  $M$  and  $m$ , the distance  $r$  between  $M$  and either of the  $m$ 's, and the angle  $\alpha$ . If either  $r$  or  $\alpha$  is altered, restoring forces come into play, and tend to resist the deformation. We may approximately describe the system in terms of a force constant,  $F$ , for the valence bonds (similar to that for a diatomic molecule) and one for the deformation of the valence angle. Two of the fundamental frequencies ( $\nu_1$  and  $\nu_2$ ) (Fig. 1) involve primarily an alteration in the length of the valence bond. In  $\nu_1$  the two atoms  $m$  move in and out together, and the vibration is symmetrical; in  $\nu_2$ , one atom  $m$  moves in toward  $M$  while the other moves away from it; the vibration is antisymmetrical. In the third frequency,  $\delta$ , the valence angle is deformed by a symmetrical oscillation of the two atoms  $m$ , the  $M$ - $m$  distance being relatively little altered by the vibration. Experience shows that this deformation frequency is considerably lower than  $\nu_1$  and  $\nu_2$ .

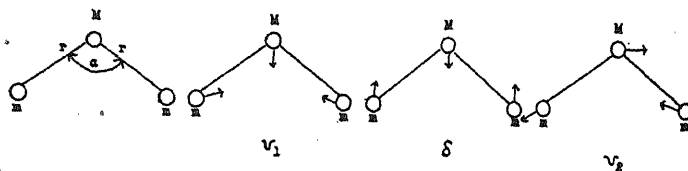


Figure 1. Model of a bent symmetrical triatomic molecule showing its fundamental types of vibration (examples:  $H_2O$ ,  $H_2S$ , or  $SO_2$ ).

The treatment of more complex molecules is fundamentally similar, but becomes increasingly elaborate as the number of atoms in the molecule increases<sup>3</sup>. Fortunately, as we have already indicated, it is found that even in complex molecules, the presence of certain groups gives rise to certain characteristic vibrational frequencies, and that such frequencies may be taken as a definite criterion of the presence of the group. This simple fact is the underlying basis of most of the following discussion<sup>4</sup>.

#### Raman Spectra<sup>2, 5, 6, 7</sup>

When a transparent diatomic substance is illuminated by light of frequency  $\nu_0$ , most of the light passes through unchanged. A small

<sup>2</sup> Kohlrausch, K. W. F., "Der Smekal-Raman-Effekt", Berlin (1931).

<sup>3</sup> In general a molecule with  $n$  atoms gives rise to  $3n - 6$  fundamental frequencies, ( $3n - 5$  for a linear molecule); but if the molecule is highly symmetrical, the number may be less. For the exact rules defining the situation see Placzek (1).

<sup>4</sup> Strictly speaking, every type of molecular vibration involves every atom in the molecule. Practically,

amount, however, is scattered in other directions by the molecules of the medium. If this scattered light is examined in a spectrograph, the original frequency  $\nu_0$  is found in high intensity. In addition, however, a series of other frequencies  $\nu_0 - \nu_1$ ,  $\nu_0 - \nu_2$  appears. The frequency shifts,  $\nu_1$ ,  $\nu_2$ , etc., are characteristic of the molecule which produces the scattering, and independent of the exciting frequency  $\nu_0$ . They are in fact vibrational frequencies of the molecule, and those commonly found among organic compounds lie in the range 150 to 4000  $\text{cm}^{-1}$ <sup>8</sup>. Some representative Raman spectra of carboxylic acids and their salts, and amino acids and their hydrochlorides, are shown in Fig. 2.

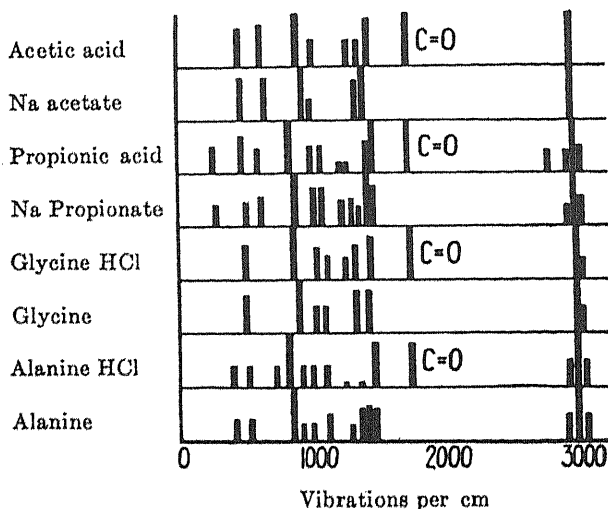


Figure 2. Raman frequencies of fatty acids and amino acids. The Raman frequencies are plotted in reciprocal wave lengths ( $\text{cm}^{-1}$ ). The height of each line is roughly proportional to its relative intensity. From Edsall, J. T., *Cold Spring Harbor Symp. Quant. Biology*, 6, 40 (1938).

In the interpretation of such vibrational spectra, one major point should be indicated here. *Only molecules or groups bound by covalent linkages give rise to Raman or infrared frequencies, in the range with which we are concerned.* That is, only such groups are capable of entering into molecular vibration; when atoms are attached only by electrovalent links, there is only general electrostatic attraction, not a definite link directed in space, and the conditions for a vibrating system do not exist. This has been shown both

<sup>8</sup> The "wave number" or reciprocal wave length (in  $\text{cm}^{-1}$ ) is now generally used as the most convenient practical measure of frequency in spectroscopy. The true frequency (in  $\text{sec}^{-1}$ ) is obtained by multiplying the wave number by  $C$ , the velocity of light in  $\text{cm/sec}$   $1 \text{ cm}^{-1} = 2.997 \times 10^{10} \text{ sec}^{-1}$ . Multiplying the frequency by  $h$ , Planck's constant,  $6.625 \times 10^{-27} \text{ erg sec}$ , gives the energy in ergs of the quantum. Multiplying the frequency by  $h \nu$ , Planck's constant,  $6.625 \times 10^{-27} \text{ erg sec}$ , gives the energy in ergs of the quantum. Multiplying the frequency by  $h \nu$ , Planck's constant,  $6.625 \times 10^{-27} \text{ erg sec}$ , gives the energy in ergs of the quantum.

experimentally<sup>9</sup> and theoretically. Thus sodium<sup>\*</sup> chloride in water gives rise to no Raman lines at all; sodium and potassium acetate give rise to an identical spectrum which is the spectrum of the acetate ion. Likewise the spectrum of the hydrochloride of an amino acid, such as glycine, is the spectrum of the amino acid cation; the isolated chloride ion plays no part in the observed vibration, except that the exact value of some of the frequencies may be modified because of the electric field produced by an ion in the neighborhood.

### The Ionization of the Carboxyl Group

The spectra in Fig. 2 show many features, but one is predominantly characteristic of the carboxyl group. This is the powerful line near  $1730\text{ cm}^{-1}$ , which is found in all un-ionized carboxylic acids, including the hydrochlorides of all the amino acids<sup>10</sup>. A similar frequency is found in all molecules containing a  $\text{C}=\text{O}$  group, such as esters, aldehydes and ketones (Table 1). In any given type of compound, its value is almost completely

TABLE 1. INFLUENCE OF VARIOUS SUBSTITUENTS ON THE  $\text{C}=\text{O}$  FREQUENCY

Substance	$\text{H}_3\text{CCOX}$	$\text{RH}_2\text{CCOX}$	$\text{R}_2\text{HCCOX}$	$\text{R}_3\text{CCOX}$	$\text{C}_6\text{H}_5\text{COX}$
Acid, $\text{X}=\text{OH}$ (anhydrous).....	1666	1652	1648	1644	1647
Acid, $\text{X}=\text{OH}$ (in water).....	1720	1719	—	—	—
Methyl ester, $\text{X}=\text{OCH}_3$ .....	1736	1735	1732	1728	1720
Ethyl ester, $\text{X}=\text{OC}_2\text{H}_5$ .....	1736	1732	1728	1724	1721
Ketone, $\text{X}=\text{CH}_3$ .....	1710	1709	1709	1702	1677
Acid chloride, $\text{X}=\text{Cl}$ .....	1798	1793	1788	1790	—
Aldehyde, $\text{X}=\text{H}$ .....	1715	1719	1719	1723	1689

Most of the data are from Kohlrausch and Pongratz [Z. phys. Chem. (B), 27, 176 (1934)]; data on the acids in water from Edsall [J. Chem. Phys., 4, 1 (1936)]. See also Hibben, J. H., Chem. Rev., 13, 345 (1933); 18, 1 (1936).

independent of the length of the attached hydrocarbon chain, R. It may, therefore, be regarded as essentially a pulsation of the  $\text{C}=\text{O}$  linkage, the rest of the molecule being relatively little involved, except through the modifying influence of immediately adjoining polar groups on the bond strength of this linkage<sup>11</sup>.

On the ionization of a fatty acid in water, this frequency completely disappears (Fig. 1). None of the sodium salts of the fatty acids or dicarboxylic acids shows a frequency near  $1700\text{ cm}^{-1}$ . This rule has been found

<sup>\*</sup> Thus  $\text{HCl}$  shows a strong Raman line near  $2880$  in the gaseous state, or at a somewhat lower frequency in solution in solvents like benzene or chloroform [West, W., and Arthur, P., J. Chem. Physics, 2, 215 (1934)]; but in aqueous solution up to 3 normal, where it is nearly completely ionized, it gives rise to no Raman frequency at all. For a further discussion of this general point, see Kohlrausch, K. W. F., Naturwiss., 22, 161, 181, 196 (1934).

<sup>10</sup> Pure anhydrous fatty acids show a lower frequency, at about  $1680$  (Kohlrausch, K. W. F., Köppl, F., and Pongratz, A., Z. physik. Chem. (B), 21, 242 (1933)). On solution of the acid in water this rises to about  $1720$  (Edsall, J. T., J. Chem. Physics, 4, 1 (1936)). This change is probably associated with the transformation of the acid in the aqueous solution. Values

to hold for a great variety of carboxylic acids, and no exceptions to it are known<sup>12</sup>.

The isoelectric monoamino-monocarboxylic acids show no frequency in this range, although their hydrochlorides show it strongly. This is definite evidence that the carboxyl group in these amino acids is ionized; in other words, that these amino acids are dipolar ions. Similarly, the hydrochlorides of the dicarboxylic amino acids, aspartic and glutamic acids, show intense Raman lines near 1740. In their monosodium salts, both carboxyl groups are ionized, and no line in this neighborhood is present.

The ionized carboxyl group not only shows no frequency near 1700, but it does give rise to one or more definite and characteristic frequencies in a lower range, near 1400. The appearance of such frequencies is particularly clearly demonstrated in malonic acid and its salts (Fig. 3).

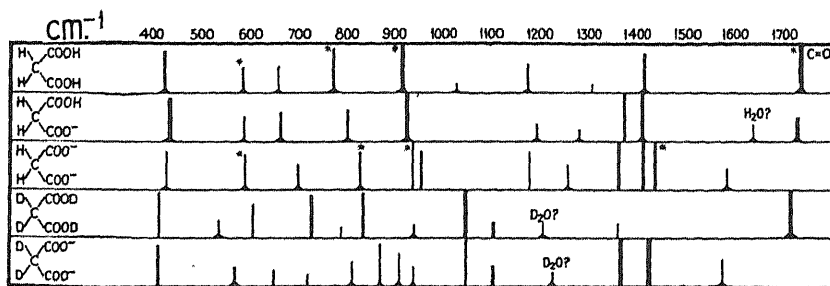


Figure 3. Raman spectra of malonic acid, malonic- $d_2$  acid- $d_2$  and their sodium salts. A broadening at the base of a line in the figure indicates that the Raman line is broad. From *Cold Spring Harbor Symp. Quant. Biol.*, 6, 40 (1938).

Malonic acid shows a powerful  $C=O$  frequency at 1738. In monosodium malonate, the  $C=O$  frequency falls slightly to 1730, and is much less intense, while a new strong line appears at 1372. In disodium malonate, the 1730 line is completely gone, and a second new line at 1439 has appeared.

The line near 1412, present in the un-ionized acid and both its salts, is unaffected by ionization. The latter line arises not from the carboxyl group, but from the  $CH_2$  group. It appears to be due to a symmetrical deformation of the  $CH_2$  group, similar to the frequency  $\delta$  for the triatomic molecule shown in Fig. 1. This line vanishes completely from the spectrum when the hydrogen in malonic acid  $CH_2(COOH)_2$  is replaced by deuterium to form  $CD_2(COOD)_2$  (Fig. 3). Instead, there appears a line of similar intensity at 1050, the frequency being depressed because of the greater mass of the deuterium atoms<sup>13</sup>. On the other hand, the vibrations

<sup>13</sup> J. H. Drenth, *J. Chem. Phys.*, 4, 508 (1937).

arising primarily from the carboxyl group are only very slightly affected by deuterium substitution, the change being of the order of 1%. The frequencies characteristic of the  $\text{CH}_2$  group are almost unaffected by ionization, but profoundly lowered by deuterium substitution; while the converse is true of the lines arising from the carboxyl group (Table 2).

These observations show two facts: (1) The ionized carboxyl group gives rise to strong Raman frequencies near  $1400\text{ cm}^{-1}$ , and (2) other frequencies in the same range may arise from  $\text{CH}_2$  or  $\text{CH}_3$  groups. The latter frequencies are greatly lowered by deuterium substitution in the  $\text{CH}_2$  or  $\text{CH}_3$  group; the former are almost unaltered. Thus the two classes of frequencies may be readily distinguished.

Extensive study indicates that a frequency near  $1400$  is as characteristic of the ionized carboxyl group as the frequency near  $1720$  is of the

TABLE 2. SOME RAMAN LINES CHARACTERISTIC OF THE  $-\text{CH}_2$ ,  $\text{COOH}$  AND  $\text{COO}^-$  GROUPS IN MALONIC ACID AND THE ANIONS OF ITS SALTS

Molecule or Ion	Frequency in $\text{cm}^{-1}$				
	$\nu_1(\text{CH}_2)$	$\nu_2(\text{CH}_2)$	$\delta(\text{CH}_2)$	$\nu(\text{COO}^-)$	$\nu(\text{C=O})$ in $\text{COOH}$
$\text{HOOC}\cdot\text{CH}_2\cdot\text{COOH}$ .....	2956 (6)	3003 (3)	1417 (4)	—	1738 (8)
$\text{HOOC}\cdot\text{CH}_2\cdot\text{COO}^-$ .....	2943 (6)	2987 (1)	1408 (5)	1372 (5)	1730 (2)
$^-\text{OOC}\cdot\text{CH}_2\cdot\text{COO}^-$ .....	2947 (6)	2984 (2)	1413 (6)	{ 1361 (6) 1439 (6)	—
$\text{DOOC}\cdot\text{CD}_2\cdot\text{COOD}$ .....	2159 (5)	2258 (3)	1050 (4)	—	1719 (7)
$^-\text{OOC}\cdot\text{CD}_2\cdot\text{COO}^-$ .....	2153 (5)	2238 (3)	1054 (4)	{ 1367 (4) 1427 (7)	—

The number in parentheses following each frequency gives a very rough idea of its intensity.  $\nu_1$  and  $\nu_2(\text{CH}_2)$  denote valence vibrations of the  $\text{CH}_2$  group, similar to  $\nu_1$  and  $\nu_2$  of water in Fig. 1.  $\delta(\text{CH}_2)$  denotes a deformation similar to  $\delta$  in Fig. 1.  $\nu(\text{COO}^-)$  denotes probably a symmetrical valence oscillation of the  $\text{COO}^-$  group (like  $\nu$  in Fig. 1) and  $\nu(\text{C=O})$  is a valence vibration of the  $\text{C=O}$  linkage in the un-ionized carboxyl group.

Data from Edsall, J. T., *J. Chem. Physics*, 5, 508 (1937).

un-ionized group. It is often, however, more difficult to recognize, since it may be masked or overlaid by a coincident line arising from a methyl or methylene group. The presence of an ionized carboxyl group in the isoelectric amino acids, however, is shown regularly by the appearance or intensification of lines near  $1400\text{ cm}^{-1}$ , while their hydrochlorides show fewer or weaker lines in this region (Table 3). The evidence on this point also thus gives definite evidence for dipolar ionic structure.

A similar intense vibration near  $1400\text{ cm}^{-1}$  is characteristic of the nitro ( $-\text{NO}_2$ ) group, which is intimately related in electronic structure to the ionized carboxyl ( $-\text{CO}_2^-$ ) group<sup>12</sup>.

#### The Ionization of the Amino Group

# SPECTROSCOPY AND DIPOLAR IONIC STRUCTURE 15

TABLE 3. EFFECT OF IONIZATION OF THE CARBOXYL GROUP ON SOME RAMAN FREQUENCIES IN THE RANGE 1300-1740  $\text{CM}^{-1}$

(Numbers in parentheses after frequency values indicate relative intensities of the lines)					
Substance	Form	COO <sup>-</sup> group (f)	COO <sup>-</sup> or CH <sub>2</sub> group	CH <sub>2</sub> or CH <sub>3</sub> group	COOH group
Acetic acid	R.COOH	1370(1)		1436(2)	1720(4b)
	R.COO <sup>-</sup>	*1347(2b)	*1413(6b)		
Propionic acid	R.COOH		1424(3)	1459(3)	1719(2b)
	R.COO <sup>-</sup>	{*1300(2) *1368(1)}	*1417(4)	1464(2)	
Malonic acid	R(COOH) <sub>2</sub>			1417(4b)	1738(8vb)
	R-COOH				
	R	*1372(5)		1408(5b)	1730(2b)
	COO <sup>-</sup>				
	R(COO <sup>-</sup> ) <sub>2</sub>	*1361(6)	*1439(6)	1413(6)	
Malonic-d <sub>2</sub> acid-d <sub>2</sub>	R(COOD) <sub>2</sub>	1360(1)		1050(4)	1719(7b)
	R(COO <sup>-</sup> ) <sub>2</sub>	*1367(4b)	*1427(7)	1054(4)	
Glycine	R.COOH	1315(2)		1436(3)	1743(3vb)
	R.COO <sup>-</sup>	*1331(3)	*1412(3)		
Alanine	R.COOH	1356(0)		1460(2b)	1738(2vb)
	R.COO <sup>-</sup>	*1358(2)	*1416(2)	1467(2)	
$\alpha$ -Amino-n-butyric acid	R.COOH	1363(1)		1450(4b)	1746(3vb)
	R.COO <sup>-</sup>	*1358(4)	*1413(3)	1459(3)	
$\alpha$ -Aminoisobutyric acid	R.COOH	1359(0)		{1452(3b) 1467(2b)}	1729( $\frac{1}{2}$ b)
				{1444(2b) 1465(3b)}	
	R.COO <sup>-</sup>	*1374(3b)	*1411(3)		
Sarcosine	R.COOH	1284(1)	1429(2)	1464(3)	1732(1b)
	R.COO <sup>-</sup>	*1320(3b)	*1408(5b)	1468(4)	
Betaine	R.COOH	1337( $\frac{1}{2}$ )	1419( $\frac{1}{2}$ )	1453(4)	1751(1b)
	R.COO <sup>-</sup>	*1335(2)	*1403(2)	1453(2)	
$\beta$ -Alanine	R.COOH	1328(2b)	1413(3b)	1472(1b)	1730(4b)
	R.COO <sup>-</sup>	1338(1)	*1410(5b)	1471( $\frac{1}{2}$ )	
$\beta$ -Aminobutyric acid	R.COOH	1380(2b)	1419(2)	1461(4)	1726(4vb)
	R.COO <sup>-</sup>	1366( $\frac{1}{2}$ )	*1408(7b)	1457(3)	
Aspartic acid	R.(COOH) <sub>2</sub>	1376( $\frac{1}{2}$ )		1418(1)	1743(4vb)
	R.(COO <sup>-</sup> ) <sub>2</sub>	{*1330(1) *1355(2)}	*1411(5b)		
Glutamic acid	R.(COOH) <sub>2</sub>	1347( $\frac{1}{2}$ )		1433(3)	1736(4vb)
	R.(COO <sup>-</sup> ) <sub>2</sub>	*1345(2b)	*1410(4b)		

The data in this table are taken from Edsall, J. T., *J. Chem. Physics*, 4: 1 (1936) and *J. Chem. Physics*, 5: 508 (1937). "b" is used to denote a broad, "vb" a very broad line. Asterisks are used to denote lines probably arising in large part from the ionized carboxyl group. These often overlap with lines arising primarily from the CH<sub>2</sub> or CH<sub>3</sub> group, and sometimes can be distinguished only by the great increase in intensity of these lines on ionization of the carboxyl group.

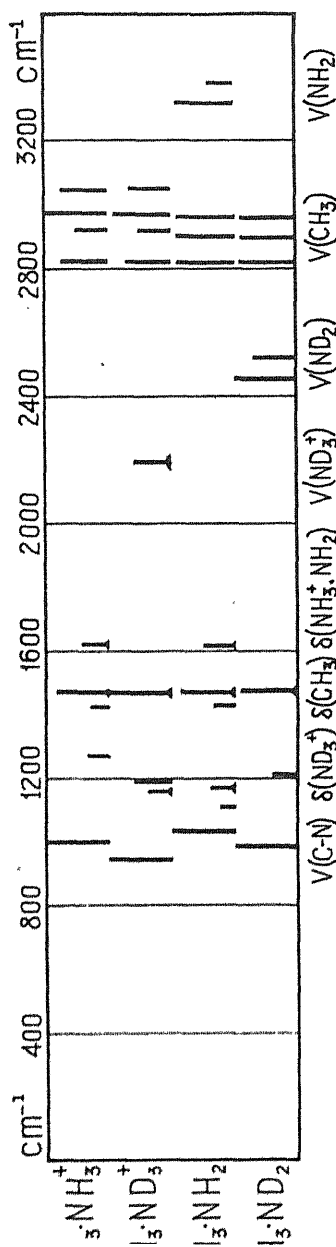


Figure 4. Raman spectra of methylamine and methylammonium ions. The symbols  $\text{V}$  and  $\delta$  indicate the region in which valence and deformation frequencies, respectively, of the various groups are found. From Edsall, J. T., and Scheinberg, J. *Chem. Phys.*, 8, 520 (1940).

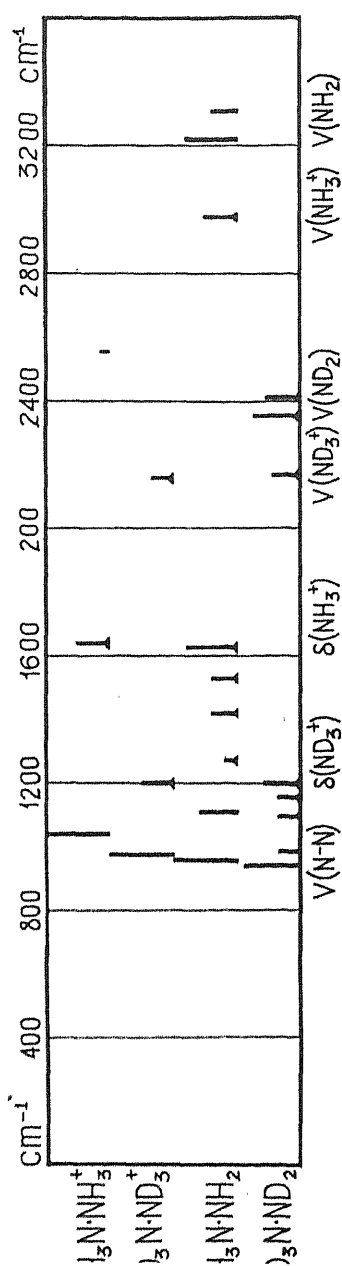


Figure 5. Raman spectra of the ions derived from hydrazine. Symbols as in Figure 4. From Edsall, J. T., and Scheinberg, H., *J. Chem. Phys.*, 8, 520 (1940).



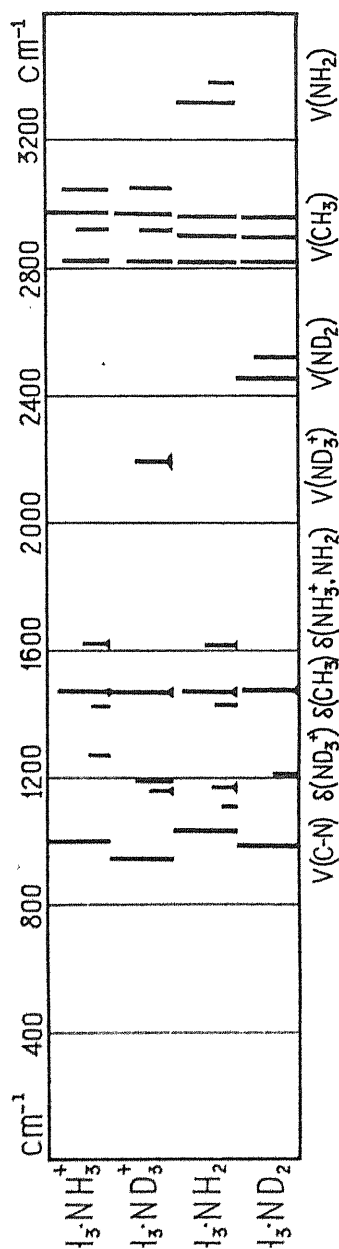


Figure 4. Raman spectra of methylamine and methylammonium ions. The symbols  $\nu$  and  $\delta$  indicate the region in which valence and deformation frequencies, respectively, of the various groups are found. From Edsall, J. T., and Scheinberg, *J. Chem. Phys.*, 8, 520 (1940).

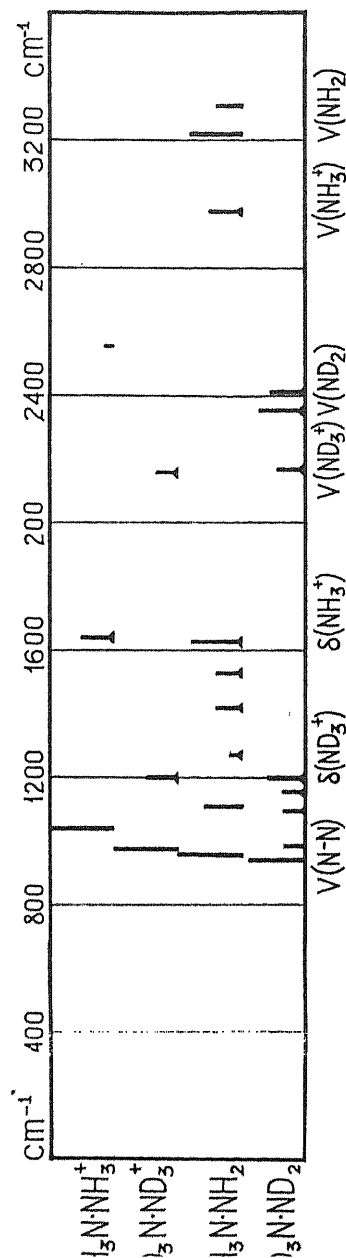


Figure 5. Raman spectra of the ions derived from hydrazine. Symbols as in Figure 4. From Edsall, J. T., and Scheinberg, H., *J. Chem. Phys.*, 8, 520 (1940).

large number of primary amines at  $3313 \pm 10 \text{ cm}^{-1}$ ; the other, less intense frequency at  $3367 \pm 15 \text{ cm}^{-1}$ <sup>14</sup>. These vibrations appear to arise primarily from the N—H valence bonds in the  $\text{NH}_2$  group. Secondary amines show only one line in this range, at about 3340, and tertiary amines show none at all. These N—H vibrations are very intense, and are practically unchanged in frequency or intensity by solution of the amine in water<sup>15</sup>. Glycine ethyl ester shows two such N—H frequencies with unusually high values: 3328 and  $3408^{16}$ . The increased value of these frequencies is probably due to the polar influence of the adjoining  $-\text{COOC}_2\text{H}_5$  group.

When the hydrogen of the  $-\text{NH}_2$  group is replaced by deuterium, its characteristic frequencies are depressed to a very marked extent. The lines at 3322 and 3382 in  $\text{CH}_3\text{NH}_2$  shift to 2450 and 2527, respectively, in  $\text{CH}_3\text{ND}_2$ <sup>17</sup>; while a set of lines near 1400 and near 2900, arising from the methyl group, is unchanged. Also hydrazine monohydrochloride ( $\text{H}_2\text{N} \cdot \text{NH}_3^+\text{Cl}^-$ ) gives a strong Raman line at 3309 (15), while  $\text{D}_2\text{N} \cdot \text{ND}_3^+\text{Cl}^-$  gives a corresponding line at  $2410^{17, 18}$  (Figs. 4 and 5).

On ionization of the amino group, these frequencies disappear. They are replaced by much broader and weaker frequencies, which in the hydrogen compounds lie in the region between 2700 and  $3200 \text{ cm}^{-1}$ , and in the deuterium compounds lie near  $2200 \text{ cm}^{-1}$ . The rather weak and diffuse character of these lines due to the charged ammonium group differentiates them strikingly from the sharper and more intense frequencies arising from the uncharged amino group, and the fact that their frequency is considerably lower indicates a weakening of the N—H bond in the alkyl-ammonium ions as compared with the uncharged amines.

Thus in  $\text{H}_2\text{N} \cdot \text{NH}_3^+\text{Cl}^-$  there is a rather weak and very broad Raman line at 2980; in  $\text{D}_2\text{N} \cdot \text{ND}_3^+\text{Cl}^-$  there is a similar line at 2175. Here again the ratio of the two frequencies is 1.371, as for the lines arising from the  $-\text{NH}_2$  and  $-\text{ND}_2$  groups. These lines, however, arise from the charged  $-\text{NH}_3^+$  and  $-\text{ND}_3^+$  groups, as is shown by a study of the methyl ammonium chlorides,  $\text{CH}_3 \cdot \text{NH}_3^+\text{Cl}^-$  and  $\text{CH}_3\text{ND}_3^+\text{Cl}^-$ <sup>17</sup> (Fig. 5). Both these compounds contain a set of four strong Raman lines—2833, 2914, 2975 and 3032—which are unaffected in position and intensity by the sub-

<sup>14</sup> Kohlrausch, K. W. F., *Monatshefte f. Chem.*, 68, 349 (1935).

<sup>15</sup> Edsall, J. T., *J. Chem. Physics*, 5, 225 (1937). The principal Raman bands of water, which lie in the range 3200–3600, overlap with these amino group frequencies. The water bands, however, are very broad and diffuse, while the N—H frequencies are very intense and narrow, so that there is no difficulty in distinguishing them.

The N—H frequencies in gaseous methylamine are higher than in the liquid, lying at 3360 and  $3470 \text{ cm}^{-1}$  (Kirby-Smith, J. S., and Bonner, L. G., *J. Chem. Physics*, 7, 880 (1939)).

The decrease in frequency on liquefaction is due to interaction between neighboring amino groups in the liquid, which lowers the strength of the N—H bond in any one molecule through the attraction exerted on

stitution of deuterium for hydrogen in the ammonium group. These lines must therefore be due to the methyl group.  $\text{CH}_3 \cdot \text{ND}_3^+ \text{Cl}^-$ , however, shows a weak diffuse line at 2180, corresponding exactly to the line at the same position in  $\text{D}_2\text{N} \cdot \text{ND}_3^+ \text{Cl}^-$ . The corresponding line in  $\text{CH}_3 \cdot \text{NH}_3^+ \text{Cl}^-$ , which should appear at about 2980, is covered up by the very strong 2975 line arising from the methyl group. Here again, as in the study of the  $\text{CH}_2$  and  $\text{COOH}$  groups in the malonic acids, the value of deuterium substitution is apparent in revealing types of vibration which would otherwise be undetected in the observed spectrum.

The Raman frequencies of the  $-\text{NH}_3^+$  group have been revealed with particular clarity by the studies of Ananthakrishnan<sup>19</sup> on the spectra of crystal powders. Crystals of hydroxylamine and hydrazine hydrochlorides show a large number of frequencies of this type, most of which cannot be found in aqueous solutions of these substances. Probably the very large number of these lines is due in part to the perturbing influence of the electrostatic forces between the ions in the crystal lattice, with resulting deformation of the  $^+\text{NH}_3\text{OH}$  and  $^+\text{H}_2\text{N} \cdot \text{NH}_3^+$  ions, and the production of many types of slightly different vibrations.

The Raman spectrum of crystalline glycine studied by the same method<sup>19, see also 12</sup> reveals no lines above  $3300 \text{ cm}^{-1}$ . There are, however, two sharp lines—at 2960 and 3000—which probably arise primarily from vibrations of the  $\text{CH}_2$  group; and three broader, fainter lines, at 2592, 2870 and  $3113 \text{ cm}^{-1}$  (the last mentioned is the strongest of the three). These correspond closely in position and appearance to the vibrations characteristic of the charged  $-\text{NH}_3^+$  group. Lines characteristic of the uncharged  $-\text{NH}_2$  group are completely absent from the spectrum. The Raman spectrum thus indicates that in glycine crystals, the amino group is positively charged, and the molecules in the crystal are, therefore, dipolar ions. This view is entirely in accord with conclusions reached from observations on melting point and crystal structure, which are discussed in later chapters.

The same conclusion is reached by the study of solutions of sodium glycinate and sodium alaninate. Their spectra show strong frequencies near 3320, characteristic of the uncharged  $-\text{NH}_2$  group<sup>15, 20</sup>. The isoelectric amino acids, on the other hand, show no indication at all of a frequency in this range. The  $-\text{NH}_3^+$  frequency near 2980 in the isoelectric amino acids is undoubtedly present, but is covered up (as in the  $\text{CH}_3\text{NH}_3^+$  ion) by the strong line of the same frequency arising from the  $-\text{CH}_2$  or  $-\text{CH}_3$  group. When the hydrogen on the  $-\text{NH}_3^+$  group in glycine is completely

replaced by deuterium, however, and the isoelectric amino acid is then dissolved in  $D_2O$ , the weak broad characteristic frequency of the charged  $-ND_3^+$  group near 2180 appears unmistakably, with no evidence of the  $-ND_2$  frequencies near 2450 and 2520<sup>21</sup>. Thus the Raman lines due to the ammonium group in isoelectric amino acids are completely in accord with the electrically charged character of a dipolar ion, and would alone furnish adequate proof of this structure.

Other applications of Raman spectra to the study of the amino acids are given in reference 7.

### Infrared Spectra

Evidence concerning the structure of the amino acids may be furnished equally well by infrared spectra, and important papers on this subject have been published by Freymann, Freymann and Rumpf<sup>22</sup>, who have shown that primary and secondary amino groups, in the uncharged form, give rise to absorption bands in the region near  $3\mu$  ( $3300\text{ cm}^{-1}$ ), and that these absorption bands vanish when the nitrogen acquires an additional proton. Similarly, the sodium salts of glycine and taurine show strong absorption bands near  $3\mu$ , which vanish in the free amino acids. This also, therefore, provides unequivocal evidence for the dipolar ionic nature of these substances. Further employment of infrared spectroscopy in this field should provide valuable information.

### Ultraviolet Spectra

Carboxyl and amino groups show strong ultraviolet absorption in the region 2200–1800 Å and probably at shorter wave-lengths. Comparison of the absorption of amino acids and their salts with that of carboxylic acids, amines and their salts has led to results which can most readily be interpreted in terms of dipolar ionic structure<sup>23</sup>. The absorption curves are somewhat more difficult of interpretation than the Raman and infrared data, but lead to the same conclusions. Thus three different types of spectroscopic study provide evidence for the existence of dipolar ions, and no evidence is found which disagrees with the hypothesis in any way.

<sup>21</sup> Unpublished observation of J. T. Edsall and H. Scheinberg.

<sup>22</sup> Freymann, R., Freymann, G., and Rumpf, P., *J. Physique et Rad.*, (7) 7, 30, 476, 506 (1936).

<sup>23</sup> Ley, H., and Arcand, B., *Z. phys. Chem.*, (B), 17, 177 (1932).

## Chapter 3

### Thermodynamics and Simple Electrostatic Theory

BY GEORGE SCATCHARD

#### Introduction

Amino acids and proteins react chemically in so many different ways, they interact physically with such large forces, and their molecules are so large, that the study of their physical chemistry leads to some fundamental problems which are not important in the study of simpler molecules, and are therefore not discussed in the more elementary books. It is well, therefore, to begin with a review of the thermodynamic treatment of equilibrium which goes back to the fundamentals, although many of the details of development must be omitted and much must be stated categorically.

The property which is most important in the study of equilibrium is called by Gibbs the chemical potential<sup>1, 2</sup>. Its magnitude depends upon the units of quantity of the component for which it is determined. We shall use the molecular weight as unit, and arbitrarily select any convenient formula which represents the composition if the true molecular weight is unknown. For many purposes this arbitrary weight serves as well as the true one, but it obviously cannot be used to determine any quantity the measurement of which would yield the true molecular weight. We call the chemical potential per mole the molal chemical potential, but we shall shorten this to potential when there is no danger of confusion with other kinds of potential or with other units of quantity.

Gibbs says: "In order to arrive as directly as possible at the most characteristic and essential laws of chemical equilibrium, we will first give our attention to a case of the simplest kind. . . . We will suppose that the case is not complicated by the action of gravity, or by any electrical influences, and that in the solid portions of the mass the pressure is the same in every direction. We will further simplify the problem by supposing that the variations of the parts of the energy and entropy which depend upon the surfaces separating heterogeneous masses are so small in comparison with

the variations of the parts of the energy and entropy which depend upon the quantities of these masses, that the former may be neglected by the side of the latter . . . all the circumstances and considerations which are here excluded will afterward be made the subject of special discussion."

However, many of his important results apply only with the restrictions given above, that is, to what we will call a restricted system. Among the results which do not apply to unrestricted systems are:

(1) The phase rule, that the variance of a system is equal to the number of components plus two minus the number of phases;

(2) The "Gibbs-Duhem" relation between the derivatives of the chemical potentials at constant temperature and pressure,

$$\sum_i n_i d\mu_i = 0; \quad (1)$$

(3) The law of "Mass Action" as applied to either physical or chemical reactions,

$$\sum_i \nu_i \mu_i = 0. \quad (2)$$

In these equations  $\mu_i$  is the molal chemical potential of the  $i$ 'th species,  $n_i$  is the number of moles of that species in the system, and  $\nu_i$  is the number of moles of the species which the reaction produces in a phase where its potential is  $\mu_i$ . The symbol  $\sum_i$  represents the summation in which  $i$  becomes in turn each species under discussion. For example, in the chemical reaction



$-a$  moles of  $A$ ,  $-b$  moles of  $B$ ,  $e$  moles of  $E$  and  $f$  moles  $F$  are produced, so

$$\sum_i \nu_i \mu_i = -a\mu_A - b\mu_B + e\mu_E + f\mu_F$$

There is nothing in these equations which requires that the units of quantity be moles, or that they even be the same for all species, for the dimensions of  $n_i \mu_i$  and  $\nu_i \mu_i$  are independent of the units of quantity, but this choice has the advantage of making the  $\nu$ 's small integers or ratios of small integers for most reactions and has other similar advantages. We shall remove some of these restrictions later, but most of our discussion will apply only to restricted systems. We shall always state when a restriction is removed, and assume that the system is restricted unless otherwise noted except when it seems desirable to emphasize particularly the restricted applicability.

The definitions of free energy,  $F$ , the work content,  $A$ , and the heat content,  $H$ , in terms of the energy,  $E$ , the entropy,  $S$ , the volume,  $V$ , the

sent the sums over the whole volume of  $PdV$  and  $(TS/V)dV$  for each infinitesimal volume. They are:

$$F = E - TS + PV \quad (3)$$

$$A = E - TS = F - PV \quad (4)$$

$$H = E + PV = F + TS \quad (5)$$

### The Chemical Potential

The relations which follow, however, apply only to restricted systems:

$$dE = TdS - PdV + \sum_i \mu_i dn_i \quad (6)$$

$$dF = dE - TdS - SdT + VdP + PdV = -SdT + VdP + \sum_i \mu_i dn_i \quad (7)$$

$$dA = -SdT - PdV + \sum_i \mu_i dn_i \quad (8)$$

$$dH = TdS + VdP + \sum_i \mu_i dn_i \quad (9)$$

The chemical potential may be defined from any of these four equations, and we will combine the four definitions in one equation

$$\mu_k = (\partial E / \partial n_k)_{S, V, n} = (\partial H / \partial n_k)_{S, P, n} = (\partial A / \partial n_k)_{T, V, n} = (\partial F / \partial n_k)_{T, P, n} \quad (10)$$

The symbol  $\partial$  is used to indicate a partial derivative, and the subscripts indicate the quantities which are held constant during the differentiation. The subscript  $n$  means that all  $n$ 's not specifically stated in the derivative are kept constant. So the first definition is an abbreviation for, "the molal potential of the  $k$ 'th component in a system is the partial derivative of the energy of that system with respect to the number of moles of the  $k$ 'th component in it when the entropy and volume of the system and the quantity in it of every component other than the  $k$ 'th is kept constant". The limitation to restricted systems implies constancy of the gravitational, electrical and magnetic potentials and the area of the surface between any two phases, and limitation of the use of the chemical potential to fluids and unstrained solids.

The partial molal free energy of Lewis and Randall<sup>2</sup> is the same as the molal chemical potential. We shall retain the simpler name of Gibbs for this case but will follow Lewis and Randall in representing any other partial quantity by the symbol for the total quantity with a bar over it and a subscript to indicate the substance whose quantity is varied, as

$$(\partial \bar{V} / \partial n_k)_{T, P, n} = \bar{V}_k$$

We shall also find considerable use for the quantity which Lewis named

in which  $a_k$  is the activity of the  $k$ 'th species and  $\mu_k^0$  is its potential in a standard state in which  $a_k$  is unity. This  $\mu_k^0$  is usually chosen as an arbitrary function of the temperature, and often also of the pressure and the solvent in condensed systems.

### The Perfect Gas

Theory and experiment agree that any gaseous mixture approaches in behavior an ideal gas mixture as the pressure is decreased. A simple way of defining an ideal gas is to say that for each component

$$\mu_k = RT \ln N_k + RT \ln P + F_{k0} \quad (12)$$

$$= RT \ln C_k + RT \ln RT + F_{k0} \quad (13)$$

in which  $N_k$  is the mole fraction of the  $k$ 'th component,  $n_k/\Sigma_i n_i$ ,  $C_k$  is its volume concentration  $n_k/V$ , and  $F_{k0}$  is a function of the temperature. The derivation of equation 13 and the commoner definitions of a perfect gas from equation 12 afford a good illustration of the application of thermodynamic methods in a simple case. In each of the following equations the first equality represents a general relationship, and the second represents the application to equation 12. The first relationship of equation 14 holds for any function which, at constant temperature and pressure, is proportional to the quantity of the phase. In the others they follow from equations 7 and 3.

$$F = \Sigma_i n_i \mu_i = \Sigma_i n_i (RT \ln N_i + RT \ln P + F_{i0}) \quad (14)$$

$$V = (\partial F / \partial P)_{T,n} = \Sigma_i n_i RT / P \quad (15)$$

$$S = -(\partial F / \partial T)_{P,n} = -\Sigma_i n_i (\ln N_i + \ln P + dF_{i0}/dT) \quad (16)$$

$$E = F + TS - PV = \Sigma_i n_i (F_{i0} - T dF_{i0}/dT - RT) \quad (17)$$

From 15 we also obtain

$$PN_k = (RT \Sigma_i n_i / V) (n_k / \Sigma_i n_i) = RTC_k$$

which shows the equivalence of 12 and 13. Equations 15, 17, and 13 are the usual definition of a perfect gas.

### Real Gases

The perfect gas laws assume that each molecule acts independently of the other molecules around it. The fact that all gases liquefy at low enough temperatures shows that there are forces of attraction between the molecules and the fact that these liquids have a finite volume shows



distance of separation that it is possible to represent their effect on the potential of any component by a series of terms, of which the first depends only upon binary encounters, the next includes ternary encounters, etc., for the number of multiple encounters in a dilute gas will decrease rapidly as the number of molecules involved increases. The number of encounters between a specified molecule of the  $i$ 'th species and a specified molecule of the  $j$ 'th species is proportional to  $1/V$ , so the number of encounters between any molecule of the  $i$ 'th species and any of the  $j$ 'th species is proportional to  $n_i n_j / V$ . The contribution of these collisions to the work content may be taken as  $\beta_{ij} n_i n_j / V$ . The total effect of binary encounters is obtained if we let  $i$  and  $j$  each represent in turn each of the species, which we shall represent by  $(\Sigma n_j)(\Sigma \beta_{ij} n_j / V)$ , or more briefly by  $\Sigma_i \beta_{ij} n_i n_j / V$ . Similarly the effect of ternary encounters will be represented by  $(\Sigma n_i)(\Sigma n_j)(\Sigma \delta_{ijk} n_k / V^2)$  or  $\Sigma_{ijk} \delta_{ijk} n_i n_j n_k / V^2$ . Then the equation for the work content becomes

$$A = \Sigma_i n_i (F_{i0} + RT \ln RT - RT + RT \ln n_i / V) + \Sigma_{ij} \beta_{ij} n_i n_j / V + \Sigma_{ijk} \delta_{ijk} n_i n_j n_k / V^2 + \dots \quad (18)$$

The  $\beta$ 's and  $\delta$ 's are functions of the temperature. We obtain the pressure by differentiating with respect to the volume

$$P = -(\partial A / \partial V)_n = RT \Sigma_i n_i / V + \Sigma_{ij} \beta_{ij} n_i n_j / V^2 + 2 \Sigma_{ijk} \delta_{ijk} n_i n_j n_k / V^3 + \dots \\ = RT \Sigma_i C_i + \Sigma_{ij} \beta_{ij} C_i C_j + 2 \Sigma_{ijk} \delta_{ijk} C_i C_j C_k + \dots \quad (19)$$

To obtain the chemical potential of the  $k$ 'th species we differentiate 18 with respect to  $n_k$ , remembering that  $i$ ,  $j$ , and  $h$  must each become  $k$  in turn.

$$\mu_k = F_{k0} + RT \ln RT + RT \ln C_k + \Sigma_i \beta_{ik} C_i + \Sigma_j \beta_{kj} C_j \\ + \Sigma_i \delta_{ijk} C_i C_j + \Sigma_{ih} \delta_{ikh} C_i C_h + \Sigma_{jh} \delta_{kjh} C_j C_h + \dots \\ = F_{k0} + RT \ln RT + RT \ln C_k + 2 \Sigma_i \beta_{ik} C_i + 3 \Sigma_{ijk} \delta_{ijk} C_i C_j + \dots \quad (20)$$

The abbreviation of the second form of this equation is possible because  $\Sigma_i = \Sigma_j = \Sigma_h$ , and there is no means of determining  $\beta_{ik}$  and  $\beta_{kj}$ , or  $\delta_{ijk}$ , independently. Thus for a two-component gas

$$P = RT(C_1 + C_2) + \beta_{11} C_1^2 + 2\beta_{12} C_1 C_2 + \beta_{22} C_2^2 + 2\delta_{111} C_1^3 + 6\delta_{112} C_1^2 C_2 \\ + 6\delta_{122} C_1 C_2^2 + 2\delta_{222} C_2^3 \quad (21) \\ \mu_1 = F_{10} + RT \ln RT + RT \ln C_1 + 2\beta_{11} C_1 + 2\beta_{12} C_2 + 3\delta_{111} C_1^2 + 6\delta_{112} C_1 C_2$$

For a van der Waals gas,  $\beta = (RTb - a)$ ,  $\delta = RTb^2$  and each of the higher coefficients is obtained by multiplying the preceding one by  $b$ . For a one component gas which forms double molecules with the association constant  $K = C_D/C_s^2$ , a function of the temperature, if  $C$  is the stoichiometric concentration,  $C_D$  the concentration of double molecules and  $C_s$  the concentration of the single molecules,  $\beta = -RTK$ , and  $\delta = 4RTK^2$ . These are the simplest physical and chemical pictures, respectively. Either could be expressed in a closed form, but the power series shows better the nature of the variation at low concentrations. It might appear from these expressions that the three types of interaction could be distinguished by the relation of the higher coefficients to the  $\beta$ 's since the higher coefficients increase the effect of repulsion, leave the effect of physical attraction unaltered and decrease the effect of chemical action. A more precise study, however, decreases the differences to such an extent that the chief distinction remaining is the difference in the sign of the first effect from the other two. The difference between the effect of  $a$  and of  $K$  does not depend upon the fact that the first is physical and the second chemical but it does depend upon the assumption in the first that each molecule of an interacting pair reacts with a third molecule as though the other were absent, and the assumption in the second that a molecule of an interacting pair does not react with a third molecule at all. The first assumption could hold only for rigid spherical molecules, and the second denies the possibility of polymers larger than double molecules. Either form is obviously too simple, and a precise treatment would yield the same result regardless of the method with which it started.

### Dilute Solutions

In liquid solutions in which one component is in great excess, it is possible to treat that component, the solvent, as a uniform medium in which the other components behave like gaseous molecules. The potential of the solvent may be calculated from those of the solutes by integrating equation 1 at constant temperature and pressure. If the concentrations of the solutes are expressed as moles per unit quantity of the solvent, the potential of that quantity of solvent minus its potential in the pure solvent has the same form as the negative of the pressure in equation 19. The molal potential which we use corresponds then to the negative of the pressure multiplied by the number of units in a mole. It is customary to take the kilogram as such a unit, and to express the concentrations as  $m_1$ , etc. The free energy corresponds to the work content of equation 18. and the sum of the chemical potentials of the solutes, each multiplied

the gas. If we reserve the subscript one for the solvent and remember that the solvent is omitted from the summations,

$$F = n_1 F_{10} + \sum_i n_i (F_{i0} - RT + RT \ln n_i / w_1 n_1) + \sum_{i,j} \beta_{ij} n_i n_j / w_1 n_1 + \sum_{i,j,h} \delta_{ijh} n_i n_j n_h / w_1^2 n_1^2 + \dots \quad (23)$$

in which  $w_1$  is one thousandth the molecular weight, so that  $m_i = n_i / w_1 n_1$ . The corresponding equations for the chemical potentials are

$$\mu_k = F_{k0} + RT \ln m_k + 2 \sum_{i,j} \beta_{ijk} m_i m_j + 3 \sum_{i,j,h} \delta_{ijkh} m_i m_j m_h + \dots \quad (24)$$

$$\mu_1 = F_{10} - w_1 (RT \sum_i m_i + \sum_{i,j} \beta_{ij} m_i m_j + 2 \sum_{i,j,h} \delta_{ijh} m_i m_j m_h + \dots) \quad (25)$$

If we consider the activities, it is natural to define the standard states such that  $\mu_1^0 = F_{10}$  and  $\mu_k^0 = F_{k0}$ . This is equivalent to defining the standard states such that the activity of the pure solvent is unity and the ratio of the activity of a solute to its concentration approaches unity as all the concentrations approach zero.

$$\ln a_k = \ln m_k + 2(\sum_{i,j} \beta_{ijk} / RT) m_i m_j + 3(\sum_{i,j,h} \delta_{ijkh} / RT) m_i m_j m_h + \dots \quad (26)$$

$$\ln a_1 = -w_1 [\sum_i m_i + (\sum_{i,j} \beta_{ij} / RT) m_i m_j + 2(\sum_{i,j,h} \delta_{ijh} / RT) m_i m_j m_h + \dots] \quad (27)$$

It is obvious that in dilute solutions the chief variation of these activities is in the first term, and that the activity coefficient,  $\gamma$ , and the osmotic coefficient,  $\varphi$ , which are defined by the first equalities in equations 28 and 29, are very useful functions.

$$\begin{aligned} \ln \gamma_k &= \ln a_k - \ln m_k = 2(\sum_{i,j} \beta_{ijk} / RT) m_i m_j + 3(\sum_{i,j,h} \gamma_{ijkh} / RT) m_i m_j m_h \\ &= 2 \sum_{i,j} B_{ijk} m_i m_j + 3 \sum_{i,j,h} D_{ijkh} m_i m_j m_h + \dots \end{aligned} \quad (28)$$

$$\begin{aligned} \varphi &= \frac{-\ln a_1}{w_1 \sum_i m_i} = 1 + \frac{\sum_{i,j} B_{ij} m_i m_j}{\sum_i m_i} + \frac{2 \sum_{i,j,h} D_{ijh} m_i m_j m_h}{\sum_i m_i} + \dots \\ &= 1 + M \sum_{i,j} B_{ij} x_i x_j + 2M^2 \sum_{i,j,h} D_{ijh} x_i x_j x_h + \dots \end{aligned} \quad (29)$$

in which  $M = \sum_i m_i$  and  $x_i = m_i / M$ .

We have already said that expansion in an integral power series is not permissible for gases composed of ions, and the same restrictions apply to solutions of ions. The Debye theory tells us that we must add terms containing the square root of the ionic strength or some equivalent quantity, to give

$$F/RT = n_1 F_{10}/RT + \sum_i n_i (F_{i0}/RT - 1 + \ln m_i + A z_i^2 \sqrt{I})$$

$$\sum_{i,j} \beta_{ij} m_i m_j / RT + (\sum_i z_i^2 / \sqrt{I}) / m_i m_j$$

in which  $z_i$  is the valence of the  $i$ 'th species and  $I = \sum_i m_i z_i^2$ . As we shall see later the Debye theory also gives the value of  $A$  in terms of the temperature and the density and dielectric constant of the solvent.

To differentiate we note that  $d\sqrt{I}/dn_k = z_k^2/2w_1n_1\sqrt{I}$  and  $d\sqrt{I}/dn_1 = -\sqrt{I}/2m_1$ .

$$\ln \gamma_k = \frac{3}{2} A z_k^2 \sqrt{I} + 2 \sum_i B_{ik} m_i + \left( 2 \sum_i C_{ik} m_i + \frac{z_k^2}{2I} \sum_{ij} C_{ij} m_i m_j \right) \sqrt{I} \\ + 3 \sum_{ij} D_{ijk} m_i m_j + \left( 3 \sum_{ij} E_{ijk} m_i m_j + \frac{z_k^2}{2I} \sum_{ijh} E_{ijh} m_i m_j m_h \right) \sqrt{I} + \dots \quad (31)$$

$$\varphi = 1 + \frac{\sqrt{I}}{2} \sum_i A_i z_i^2 x_i + M \sum_{ij} x_i x_j \left( B_{ij} + \frac{3}{2} C_{ij} \sqrt{I} \right) \\ + M^2 \sum_{ijh} x_i x_j x_h \left( 2 D_{ijh} + \frac{5}{2} E_{ijh} \sqrt{I} \right) + \dots \quad (32)$$

Although equation 31 is written for a single ion species, it may be used only in a sum of ionic potentials,  $\sum_{j \neq i} \mu_j$ , so chosen that the sum of the valences,  $\sum_{j \neq i} z_j$ , is zero. If there are any ion species involved, there must be at least two. In the simplest case there are only two, and there are the same number of equivalents of each. If they have charges of opposite sign, both are added together; if they have charges of the same sign, one is removed as the other added, always in equivalent quantities. The reason for this restriction is that the net charge on the solution must remain zero. Otherwise there would be an amount of electrical work very large compared to the chemical potential difference and depending on the size of the system and its relation to neighboring systems. We will discuss this limitation more fully later.

The quantity  $I$  is twice the ionic strength of Lewis and Randall, and differs from the ionality,  $\Gamma$ , in that the concentrations are expressed in moles per kilogram of solvent in  $I$ , and as moles per liter of solution in  $\Gamma$ . It is worth noting that the dependence upon the square of the valence and upon the square root of the concentration were both first discovered empirically by Mellanby<sup>3</sup> in his study of the solubilities of globulins in salt solutions.

### Ideal Solutions

The above treatment is not adequate for very concentrated solutions or for the transfer from one solvent to another, but we can utilize the fact that some of the properties of perfect gas mixtures persist in certain

$nRT$  and even to the liquid state. For example the volume increase,  $\Delta V$ , and the heat absorbed,  $\Delta H$ , on mixing at constant temperature and pressure, are often practically zero. It is usually agreed that the third characteristic is that, at constant temperature and pressure, the potential of any component is given by the relation

$$\mu_k = \mu_k^0 + RT \ln N_k \quad (33)$$

which is equivalent to either of the following

$$a_k = N_k \quad (?)$$

$$F = \sum_i n_i (\mu_i^0 + RT \ln N_i)$$

We apply the term ideal solution to a mixture which has these three characteristics.

The equivalent of equation 33 in terms of vapor pressures was first obtained empirically by Raoult. Our present theoretical methods are not adapted to settle the question of what the third relation should be in a condensed system in which the first two relations hold. We may obtain considerable justification for equation 33, though no proof, by considering a gas under higher and higher pressures. At very low pressures a gas mixture is ideal. If the volumes are additive at all pressures,  $\int \Delta V dP$  is also zero and equation 33 will hold at all pressures. However, we know that when  $\Delta V$  is not zero, it often first increases with increasing pressure and then decreases. It is conceivable that gases mix with no volume change at high pressures, but with considerable volume change at lower pressures, so that  $\Delta F - \Delta F_0 = \int \Delta V dP \neq 0$  though  $\Delta V = 0$ . If the heat of mixing is zero we know that

$$\Delta H = \int [\Delta V - T(\partial \Delta V / \partial T)_{P,n}] dP = 0$$

If this relation is to hold at all pressures, it follows that  $\Delta V = T(\partial \Delta V / \partial T)_{P,n}$  or that  $\Delta V$  is proportional to the absolute temperature. Except when  $\Delta V$  is zero such a relation is contrary both to measurements and to the theory of gases, which agree that  $\Delta V$  decreases with increasing temperature. Neither measurements, nor theory are complete enough to prove that  $\Delta V$  may not change sign with pressure in such a way that  $\Delta H$ , like  $\Delta V$ , may increase to a maximum and then decrease, perhaps to zero, but such behavior does appear very unlikely. It is possible to continue this

to a temperature below the critical temperature at constant pressure, and the pressure is then decreased at constant temperature to any value above the vapor pressure at that temperature. The conditions for the holding of equation 33 throughout this process are that the heat of mixing shall be zero at all temperatures from the high to the low and at the high pressure, and that the volume change shall be zero at the low temperature and at all pressures from the high to the low.

From time to time it is contended that, instead of equation 33 to define ideal solutions, we should have

$$\mu_k - \mu_k^0 = RT \ln C_k/C_k^0 \quad (36)$$

which is equivalent to  $a_k = C_k/C_k^0$ . The criterion that this shall hold for a gas is that the molal volume shall be independent of the composition for all pressures smaller than that at which the relation is to be applied. This means not only that the volume must be additive, but also that the molal volumes of all the components must be the same. In fact this same condition is necessary if the relation is to hold for all the components at any one pressure, for it leads to

$$F = \sum_i n_i (\mu_i^0 + RT \ln C_i/C_i^0) \quad (37)$$

and the condition that  $(\partial F/\partial n_k)_{T,P,n} = \mu_k$ , if  $F$  and  $\mu_k$  are given by equations 37 and 36, is that  $\sum_i C_i$  is independent of the composition. The distinctions between equations 33 and 36 become particularly important when small molecules are replaced by large. Most organic molecules are several times as large, and most protein molecules are several thousand times as large, as a water molecule. So the error of using equation 36 might be particularly large in protein solutions.

The activity coefficient and the osmotic coefficient minus one of an ideal solution, as defined in equations 28 and 29, are not zero but

$$\ln \gamma_1 = -\ln (1 + w_1 M) = -w_1 M + (w_1 M)^2/2 - (w_1 M)^3/3 + \dots \quad (38)$$

$$\varphi_T = [\ln (1 + w_1 M)]/w_1 M = 1 - w_1 M/2 + (w_1 M)^2/3 - (w_1 M)^3/4 + \dots \quad (39)$$

For water,  $w_1$  is 0.018, so the activity coefficient is reduced about two per cent for each mole of solute per kilogram of water. The effect is not large, but allowance should be made for the fact that the coefficients of an ideal solution are  $B_T = -w_1/2$  and  $D_T = w_1^2/6$ .

### Chemical Equilibria

The most important uses of chemical thermodynamics are the applications of equation 2. It applies not only to chemical reactions but also

are in equilibrium with respect to the transfer of that substance. In electrolyte solutions there cannot be equilibrium with respect to the transfer of a single ion species, but at equilibrium  $\sum_i \nu_i \mu_i = 0$  if  $\sum_i \nu_i z_i = 0$ . Also in the case of a compound which separates without forming a solid solution, the chemical potential of either component need not be the same as in the solution, but only the sum  $\sum_i \nu_i \mu_i$  need be the same if each  $\nu$  represents the number of moles of the corresponding component in the compound. Thus the same solid may be in equilibrium with either of two solutions which are not in equilibrium with each other, for either component may be transferred alone from one solution to the other.

For chemical reactions it is convenient to combine equation 11 with equation 2 to give

$$\sum_i \nu_i (\mu_i^0 + RT \ln a_i) = 0$$

or

$$\sum_i \nu_i \ln a_i = -\sum_i \nu_i \mu_i^0 / RT = \ln K \quad (40)$$

and for solutes it is convenient to make the further substitution

$$\sum_i \nu_i \ln m_i = \ln K - \sum_i \nu_i \ln \gamma_i = \ln K' \quad (41)$$

$K$  is the "thermodynamic mass action constant". It is a function of the temperature. It is also a function of the pressure and the solvent as used in equation 41, and it may be so in 40.  $K'$  is a function also of the concentration of each solute. It is the familiar mass action constant in terms of molalities. Similar eliminations of the deviations from ideal solution laws or perfect gas laws give the mass action constant in terms of mole fractions, partial pressures or volume concentrations. A convenient general method of treating the components, if there are no gases under high pressure, is molality times activity coefficient for each solute, vapor pressure divided by that of the pure substance for the solvent, partial pressure for each gas, and unity for each solid or insoluble liquid.

One of the important achievements of the Gibbs method is to break down the study of complicated equilibria into the study of the chemical potentials of the individual components. This procedure depends upon the fact that the chemical potential of a component depends only upon the medium and the conditions, but not at all upon the equilibrium which is being considered. Therefore a table of potentials at unit activity at a standard temperature is sufficient to determine the equilibrium of every chemical or physical reaction between any of the substances whose potentials are included at that temperature and at such low concentrations of gases and solutes that they may be considered perfect.

ion may be given arbitrarily the value of zero potential at unit activity,  $\mu_g^0 = 0$ . Such a tabulation was initiated by G. N. Lewis under the name of standard free energies of formation<sup>2</sup>. The variation of the constants with the temperature may be determined if the partial heat contents are also tabulated. If these changes are carried over a large temperature range, the heat contents must be known over the whole range. It is important to treat the standard free energy of a gas or a solute, not as the value at unit activity, but as the value at an extremely low pressure  $P^*$  minus  $RT \ln P^*$  or the value at an extremely small concentration  $m^*$  minus  $RT \ln m^*$ . Therefore it is the partial heat content at low pressure or concentration which must be used.

The study of the variation of the ordinary equilibrium constants with pressure or with the concentration of solutes may also be split into the variation of the potentials of the individual components, and may be further reduced to the study of the variation of the activity coefficients, if we use the term here to include  $a_k/N_k$ ,  $a_k/P_k$  and  $a_k/C_k$  as well as  $a_k/m_k$ . The changes of activity coefficient with temperature may be determined from the partial heat contents relative to those in the standard state.

It is often more convenient to find the variation of  $F/T$  than that of  $F$  itself. From equations 5 and 7 and elementary calculus we have

$$(\partial F / \partial T)_{P,n} = -S = (F - H)/T \quad (42)$$

$$[\partial (F/T) / \partial T]_{P,n} = -H/T^2 \quad (43)$$

$$[\partial (F/T) / \partial (1/T)]_{P,n} = H \quad (44)$$

These equations apply equally to the partial quantities  $\mu_k$ ,  $\bar{H}_k$  and  $\bar{S}_k$ .

### ✓ The Measurement of Chemical Potentials.

Most of the substances which interest us here are far too complicated to be included in a table of standard free energies, but these very complications magnify the change of their activity coefficients with concentration of themselves or other solutes, and the effect of changes in their concentrations upon the activity coefficients of other components. These changes are determined just as those for simple substances by measuring the change in chemical potential with changing medium.

1. For a gas, the change in free energy with pressure, as determined by the equation  $(\partial F / \partial P)_{T,n}$ , is large, and the chemical potentials may be determined by differentiation. This method is used mostly for one component gases, and mixtures are studied by measuring the change in concentration or partial pressure,  $PN_k$ , when other gases are added, but the



tential of the solid or liquid caused by the change in pressure or by the solubility of the other gases may be calculated if necessary.

The equating of the chemical potential in two phases which are in equilibrium with respect to the transfer of the corresponding component is the almost universal method for the study of condensed systems. If the comparison is with a gas, it may be made at constant temperature by measuring the partial pressures in the vapors in equilibrium with the two phases and calculating the difference in chemical potentials from the behavior of the vapor. For many purposes it is sufficient to assume that the vapor is a perfect gas, but more refined methods are usually available.

The comparison with a gas may also be made at constant pressure, and the simplest case is that in which the other components are non-volatile. At the boiling point of the solution, the potential of the volatile component is the same in the gas as in the solution, so the difference between the potential in the solution and in the pure liquid is the same as that between the potential in the gas and in the liquid. This difference is zero at the boiling point of the pure liquid, and may be calculated at any other temperature from the properties of the pure substance.

$$\mu_k - \mu_k^0 = \mu_g - \mu_k^0 = T \int_{T_k}^T d(\mu_g/T - \mu_k^0/T) = T \int_{T_k}^T (\Delta H/n) d(1/T) \quad (45)$$

in which  $\Delta H/n$  is the difference in heat content per mole in the vapor and in the liquid when each is at the boiling pressure. At the boiling point, it is, therefore, the heat of evaporation, and its change with temperature is the difference in constant pressure heat capacities of the vapor and the liquid. Equation 45 is exact, and the accuracy of its integration depends only upon the knowledge of the heat capacity difference. For many purposes this difference may be neglected to give

$$\mu_k - \mu_k^0 = (\Delta H/n)(1 - T/T_k) \quad (46)$$

$$\ln a_k = \frac{\Delta H/n}{R} \left( \frac{1}{T} - \frac{1}{T_k} \right) = \frac{\Delta H/n}{RTT_k} (T_k - T) \quad (47)$$

None of the special characteristics of a gas are involved in the derivation of equations 45, 46, and 47. Equation 45 applies equally to freezing if  $\Delta H/n$  is the difference in heat content per mole between the solid and the liquid (which is negative) and if the solid phase contains only the  $k$ 'th component, and equations 46 and 47 apply if also the difference between the heat capacities of the solid and the liquid may be ignored. The facts that the boiling point is elevated and the freezing point depressed depend

boiling point elevation and freezing point depression are derived from equation 47 by making a series of approximations, the first of which is to replace  $a_1$  by  $N_1$ .

There are probably theoretical reasons for believing that no substance has zero vapor pressure or zero solubility in any phase, that is that no substance can have zero concentration in a phase which is in equilibrium with another phase in which its concentration is finite. There has been much discussion of this point in the literature, but no improvement over the treatment of Gibbs. He notes that in case the concentration is zero in one of the phases in equilibrium, we cannot say that the potential in that phase is equal to that in a phase in which the concentration is finite, but only that it is not less in the first phase than in the second, and Gibbs uses the symbol  $\geq$  throughout to take care of this possibility. However, he notes that for ordinary substances the potential approaches minus infinity as the concentration goes to zero, as is shown by equations 20 and 24. So the concentration can be zero only if the behavior of a substance is very different from that in phases in which it exists at higher, though perhaps still very small, concentrations. For practical purposes, however, we may consider a substance as absent from a phase if it is not present in large enough quantities to be detected itself by the methods we are using, or to change the potential of any other component beyond the precision of our measuring it. There are certainly very many cases in which the concentrations are below this practical limit.

3. We are not able to study either the vapor pressure, the boiling point or the freezing point of an amino acid or protein, and we are limited to the study of the effect of changing the medium upon the solubility, that is, upon the concentration at which the chemical potential is equal to that of the solid. It is possible, however, to use equation 1 to calculate the chemical potentials of these solutes from the potentials of the solvent measured by its vapor pressure, boiling point elevation, freezing point depression or osmotic pressure. The last method will be discussed in detail more conveniently later. Equation 1 is not convenient to use directly, for the chemical potential goes to minus infinity at zero concentration. This difficulty may be avoided by subtracting the ideal solution term before integrating. If the solution has more than two components, it is necessary to keep the ratios of the quantities of all but one constant, and thus treat them as a complex component. We may write equation 11 as

$$\mu_k = \mu_k^0 + RT \ln a_k$$

Dividing by  $RT(n_1 + n_2)$ , we may write equation 1 for a two component system as

$$N_1[d \ln N_1 + d \ln (a_1/N_1)] + N_2[d \ln N_2 + d \ln (a_2/N_2)] = 0$$

Since  $N_1 + N_2 = 1$ ,  $N_1 d \ln N_1 + N_2 d \ln N_2 = 0$ , so

$$N_1 d \ln (a_1/N_1) + N_2 d \ln (a_2/N_2) = 0 \quad (48)$$

or

$$\ln (a_2/N_2) = - \int \frac{N_1}{N_2} d \ln (a_1/N_1) \quad (49)$$

With the activity coefficient and osmotic coefficient as defined in equations 28 and 29, we use the relation

$$N_2 = 1 - N_1 = w_1 M / (1 + w_1 M)$$

to obtain

$$d \ln \gamma = d \ln (a_2/M) = - \frac{d \ln a_1}{w_1 M} - \ln M$$

Integrating by parts, choosing the constant so that  $\ln \gamma = 0$  when  $M = 0$ , and replacing  $-(\ln a_1)/w_1 M$  by  $\varphi$ , we obtain

$$\ln \gamma = \varphi - 1 + \int_0^M (\varphi - 1) d \ln m = \varphi - 1 + \int_0^M \frac{(\varphi - 1)}{m} dm \quad (50)$$

The uncertainty in  $\ln \gamma$  is approximately twice the uncertainty in  $\varphi$  plus the uncertainty in the integration from zero concentration to that of the first measurements. Theory is invoked to reduce the latter uncertainty, and it is done conveniently by assuming the validity of equation 29 or 32 in very dilute solutions.

✓ Measurements of boiling point elevation or freezing point depression give the osmotic coefficient at the boiling point or freezing point of the solution, so the temperature varies with the concentration. The values should be calculated to a constant temperature. Moreover, it is often important to know the values of  $\ln \gamma$  at temperatures other than those at which they have been measured. From equation 44 we obtain

$$d \ln \gamma_k / d(1/T) = (\bar{H}_k - \bar{H}_k^{01}) / R \quad (51)$$

$$d\varphi / d(1/T) = -(\bar{H}_1 - \bar{H}_1^0) / RM \quad (52)$$

in which  $\bar{H}^0$  and  $\bar{H}^{01}$  are the values of  $\bar{H}_1$  and  $\bar{H}_k$  when  $M$  is zero and 1 is

heat contents and not in partial quantities. The heat of dilution to zero concentration  $(H - H^*)$  is given by

$$(H - H^*) = \Sigma_i n_i (\bar{H}_i - \bar{H}_i^{01}) + n_1 (\bar{H}_1 - \bar{H}_1^0) \quad (53)$$

The corresponding free energy is given by

$$\begin{aligned} (F - F^*) &= \Sigma_i n_i (\bar{F}_i - \bar{F}_i^{01}) + n_1 (\bar{F}_1 - \bar{F}_1^0) \\ &= RT [\Sigma_i n_i (\ln m_i + \ln \gamma_i - \phi)] \end{aligned} \quad (54)$$

we may rewrite equation 30 as

$$\begin{aligned} (F - F^*)/RT &= \Sigma_i n_i (\ln m_i - 1 + A z_i^2 \sqrt{I}) + \Sigma_{ij} n_i m_j (B_{ij} + C_{ij} \sqrt{I}) \\ &\quad + \Sigma_{ijk} n_i m_j m_k (D_{ijk} + E_{ijk} \sqrt{I}) + \dots \end{aligned} \quad (55)$$

If equation 55 is to hold over a range of temperature it is necessary that

$$\begin{aligned} (H - H^*)/R &= \Sigma_i n_i z_i^2 a \sqrt{I} + \Sigma_{ij} n_i m_j (b_{ij} + c_{ij} \sqrt{I}) \\ &\quad + \Sigma_{ijk} n_i m_j m_k (d_{ijk} + e_{ijk} \sqrt{I}) \dots \end{aligned} \quad (56)$$

and that the heat capacity and its temperature derivatives be represented by  $a' = da/dT$ ,  $a'' = da'/dT$ , and similarly, for the other parameters and the higher coefficients. It is obvious that  $a = dA/d(1/T)$ , etc. to  $e_{ijk} = dE_{ijk}/d(1/T)$ . Therefore

$$\begin{aligned} d \ln \gamma_k / d(1/T) &= \frac{3}{2} z_k^2 a \sqrt{I} + 2 \Sigma_i b_{ik} m_i + \left( 2 \Sigma_i c_{ik} m_i + \frac{z_k^2}{2I} \Sigma_{ij} c_{ij} m_i m_j \right) \sqrt{I} \\ &\quad + 3 \Sigma_{ij} d_{ijk} m_i m_j + \left( 3 \Sigma_{ij} e_{ijk} m_i m_j + \frac{z_k^2}{2I} \Sigma_{ijk} e_{ijk} m_i m_j m_k \right) \sqrt{I} + \dots \end{aligned} \quad (57)$$

and

$$\begin{aligned} d\phi/d(1/T) &= \frac{a \sqrt{I}}{2} \Sigma_i z_i^2 x_i + M \Sigma_{ij} x_i x_j \left( b_{ij} + \frac{3}{2} c_{ij} \sqrt{I} \right) \\ &\quad + M^2 (\Sigma_{ijk} x_i x_j x_k \left( 2d_{ijk} + \frac{5}{2} e_{ijk} \sqrt{I} \right) \end{aligned} \quad (58)$$

✓ The partial heats may, of course, be determined graphically, but some equivalent of these analytical expressions must be used in very dilute solutions, particularly for the  $a$ 's, which may be determined by the Debye theory.

For a two-component system, in which  $H = n_1 \bar{H}_1 + n_2 \bar{H}_2$  and  $H^* = n_1 \bar{H}_1^0 + n_2 \bar{H}_2^{01}$ ,  $(H - n_1 \bar{H}_1^0)/n_2$  is usually known as the apparent molal heat content of the second component and  $(H - H^*)/n_2$  is known as the

relative apparent molal heat content, and is often given the symbol  $\Phi_{H_2}$ . Similarly  $(C_p - C_p^*)/n_2$  is called the relative apparent molal heat capacity at constant pressure of the second component, and is represented by  $\Phi_{C_p,2}$ . For the heat content it is the difference between the apparent molal heat content at two different finite compositions which is determined directly from a heat of dilution measurement. For the heat capacity, as for the volume, compressibility and thermal expansion, a single value of the apparent molal heat capacity is determined. For the free energy on the other hand, it is the difference in partial molal free energy, or chemical potential, of one component in two states at the same temperature which is determined:

### Unrestricted Systems

Before discussing unrestricted systems we will give a more general and more complete statement of the criteria of equilibrium. The general treatment would be to add to equation 6, and therefore to equations 7, 8 and 9, a term  $dE'$  for each additional process by which the energy of the system can be increased. We will write the equation for the inclusion of three typical processes which are the most important to us.

$$dE = TdS - PdV + \sum \mu_i dn_i + GdW + \mathcal{E}de + \sigma d\mathcal{Q} \quad (59)$$

in which  $G$  is the gravitational potential,  $W$  is mass,  $\mathcal{E}$  is electrical potential difference, while  $e$  is the electrical charge,  $\sigma$  is surface tension, and  $\mathcal{Q}$  is surface area. It is understood that there are as many electrical potential differences as there are pairs of electrodes to measure them, and as many surface tensions as there are kinds of interfaces between phases. Gibbs' criterion of equilibrium is that  $dE = 0$  for any possible variation at constant entropy, volume and quantity of matter. His method is to sum equation 59 for each of homogeneous parts of the system, each of which may be only infinitesimal in extent, and to determine the conditions which are necessary and sufficient for the criterion.

Since heat may flow without other change, the temperature must be uniform throughout. Otherwise the sums of  $dS$  and  $TdS$  could not each be zero. In a restricted system the pressure is uniform throughout for the volumes of different parts may be changed without any change in chemical potentials by mass transfer of matter without any change in composition. However, if there is a diaphragm through which one or more components cannot pass, the pressure may be different on the two sides of the diaphragm though uniform on either side. In an unrestricted system, the mass transfer of matter through a potential gradient may

constant potential. It follows directly from the criterion of equilibrium and equation 59 that

$$\sum_i \mu_i dn_i + GdW + \mathcal{E}de + \sigma d\mathcal{Q} = 0 \quad (60)$$

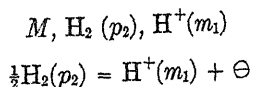
for any possible change. For a restricted system this reduces to  $\sum_i \mu_i dn_i = 0$ . For a system which is restricted except for the fact that electricity may be circulated through a single pair of electrodes,

$$\sum_i \nu_i \mu_i + \nu_e \mathcal{E} \mathcal{F} = 0 \quad (61)$$

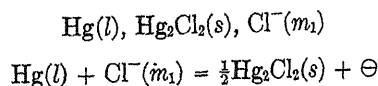
in which  $\nu_e$  is the number of equivalents of electricity transferred through the system,  $\mathcal{E}$  is the electromotive force, or electrical potential difference, and  $\mathcal{F}$  is the value of the Faraday. Since the work necessary to transfer electricity from one medium to another depends upon the magnitude of the charged body transferred, the difference in electrical potential between two different media has no exact meaning, but the electromotive force must be measured between two specimens of the same material. In practice this is usually between two pieces of copper. Since the electromotive force would not be altered by the insertion of any number of metals between two pieces of copper if the temperature were uniform throughout, the description of a cell is usually taken for granted beyond the first metal connection. Thus the simplest cell must include two metal electrodes with a solution between them. A more complicated cell may have more solutions, with or without pairs of electrodes between the solutions. It is customary in this country to write the successive phases with commas between, and to call the electromotive force positive if positive electricity would pass through the cell from left to right were the cell short-circuited to permit the cell reaction to take place with no opposing electromotive force.

The essential of an electrode reaction is that the production of an electron in the electrode is accompanied by a change of state, and that the removal of an electron is accompanied by a reversal of the change. At least one reactant must be an ion and therefore in the solution. The other reactants may be in the solution, part of the electrode, or in a third phase. It is not necessary that there be but a single reaction, provided that the reactions occur at equilibrium and reversibly. The study of equilibrium is not concerned with the mechanism of the reaction, but only with the net change in state. The conditions under which the commoner electrodes react reversibly are now pretty well known.

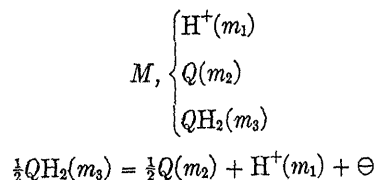
at which the hydrogen ion reacts and the calomel electrode. We write the hydrogen electrode and the electrode reaction as



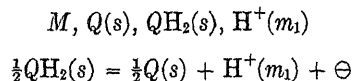
in which  $M$  represents an inert metal electrode.  $\ominus$  represents an electron, and the symbols in parenthesis represent the state of aggregation: solid ( $s$ ), liquid ( $l$ ), gas ( $g$ ), or the pressure of a gas ( $p_2$ ), or concentration of a solute ( $m_1$ ). The calomel electrode is written



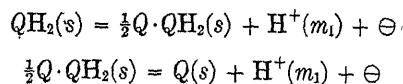
If we represent quinone and hydroquinone by  $Q$  and  $QH_2$ , the corresponding electrode is



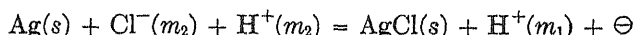
The change in free energy depends upon the chemical potentials of both the quinone and the hydroquinone as well as that of the hydrogen ion. Even if the concentration of the first two are known, the ratios of their activity coefficients are also involved, and this ratio may vary with changing medium. This difficulty may be avoided by saturating the electrode solution with the quinone and hydroquinone to give



in which the chemical potentials of the solids are constant. This may be done only with substituted quinones. For quinone itself, quinhydrone ( $Q \cdot QH_2$ ) is so insoluble that a solution saturated with both quinone and hydroquinone is not stable. However, saturation with quinhydrone and either quinone or hydroquinone, serves the same purpose, for the reactions are



tigation and one of fixed concentration of hydrogen ion and of the ion which reacts at the actual electrode. A common one is



Since the solution in which the composition of  $\text{H}^+$  and  $\text{Cl}^-$  is  $m_2$  is kept constant, the only variable is  $m_1$ , the concentration of the hydrogen ion in the solution under investigation.

### Liquid-Junction Potentials

If the cell contains but a single solution, the total change in state is that of the electrode processes, for the transfer of ions which carries the electricity from one electrode to the other is a transfer between two points in the same medium, that is, where their chemical potentials are the same. If there are two solutions without a pair of electrodes between them, however, there is a change in free energy due to the transfer of material through the boundary. It is convenient to define a transference number which differs from that in general use. We shall define the transference number of the  $i$ 'th species,  $t_i$ , as the number of moles of that species carried through the solution in the direction of positive current when one equivalent of electricity is passed through. The usual transference number is this number multiplied by the valence, which is taken negative for negative ions. Thus it is the sum of the products of transference number and valence, and not the sum of the transference numbers themselves, which equal unity. More important is the fact that neutral species have transference numbers although they do not contribute to the transport of electricity. They may, however, contribute largely to the change in free energy. The change in state for a cell containing one electrode in the solution represented by primes and one in the solution represented by double primes is

$$\Delta F = \sum_i \nu'_i \mu'_i + \nu_e \int \sum_i t_i d\mu_i + \sum_i \nu''_i \mu''_i \quad (62)$$

in which the first and third terms result from the electrode reactions, and the  $\nu$ 's represent the number of moles which are produced at the corresponding electrode, the second term arises from the transference, and  $\nu_e$  is the number of equivalents of electricity passed through the cell.

If the cell is to have a reproducible electromotive force, it is essential that there shall be a single measure of progress from the first solution to the second, which measure we may call  $x$ , such that each transference



regardless of the path. The mathematical term for such a relation is cylindrical symmetry, which is therefore seen to have no necessary dependence upon the geometry of the transition boundary. Even when the arrangement is such that  $x$  is to be measured along the axis of a right cylinder, the distance along this axis is a very inconvenient measure. If the solvent is the same in both solutions, the simplest measure is the quantity of solvent, since this automatically makes the transference number of the solvent zero throughout. This method makes the "Hittorf transference numbers" the ones to be used, except for our convention about sign. If the distance were chosen, the "true transference numbers" should be used, including one for the solvent. The same result would be obtained in the end, but with greater difficulty.

It is convenient to split the liquid junction term into three parts by dividing each chemical potential into  $\mu_i^0 + RT \ln m_i + RT \ln \gamma_i$  and each transference number into two parts  $t_i = t_i^0 + t_i' = m_i u_i^0 / \sum_j m_j u_j^0 z_j + t_i'$  (63) in which  $u_i^0$  and  $u_j^0$  are mobilities assumed independent of the composition, and most conveniently chosen for each species as its mobility in the end solution in which its concentration is largest. Then

$$\begin{aligned} \int \Sigma_i t_i d\mu_i / RT &= \int \Sigma_i t_i^0 d \ln m_i + \int \Sigma_i t_i' d \ln m_i + \int \Sigma_i t_i d \ln \gamma_i \\ &= \int \Sigma_i u_i^0 dm_i / \Sigma_j m_j u_j^0 z_j + \int \Sigma_i t_i' d \ln m_i + \int \Sigma_i t_i d \ln \gamma_i \end{aligned} \quad (64)$$

The second and third integrals are correcting terms which would vanish if the ions were ideal solutes with constant mobilities. For the cases which have been calculated, the sum is much smaller than either one alone. Therefore the first term is the most important by far. If there is a single solute with but two species of ions, the fact that the solution must be electrically neutral throughout the boundary makes  $m_1 z_1 = -m_2 z_2$  throughout, and the integral becomes  $(t_1 + t_2) \ln (m''/m')$ . In this case the other integrals are also determined by the end solutions only.

If there are more than two ion species, the integral depends upon the compositions of all the boundary layers. If the boundary is made by stirring together the two solutions with no diffusion, the composition of each layer may be represented as  $x$  parts of the first and  $(1-x)$  parts of the second, and  $x$  varies continuously from zero to unity. This is called a "mixture boundary". Henderson<sup>4,5</sup> studied this type of junction and obtained the integral

$$\int \Sigma_i u_i^0 dm_i \quad \Sigma_i (m_i'' - m_i') u_i^0 \quad \Sigma_i m_i'' u_i^0 z_i$$

Suppose that the second solution is the solution under investigation, and the first may be varied arbitrarily to make the value of the integral as small as possible. The integral will vanish as  $(\Sigma_i m'_i u_i^0 z_i) / (\Sigma_i m''_i u_i^0 z_i)$  approaches infinity if  $\Sigma_i m'_i u_i^0$  is zero. Potassium chloride is usually used for salt bridges because the mobilities of its two ions are nearly the same, or  $\Sigma_i m_i u_i^0$  is very small, and a saturated solution is used to get a large conductance,  $\Sigma_i m'_i u_i^0 z_i$ . The efficiency of the bridge in decreasing this integral will lessen as  $\Sigma_i m''_i u_i^0$  increases. In practice this means that the efficiency is small in aqueous solution if there is much of either hydrogen ion or hydroxyl ion present, for both of these ions have mobilities very large relative to those of other ions.

The Clark<sup>6</sup> type of junction and the flowing junction<sup>7, 8, 9</sup> are probably very nearly mixture boundaries. Diffusion has very little chance to change the flowing junction because it is renewed so rapidly, and it has little chance to change the Clark junction during the course of an ordinary measurement because the junction is so broad. Planck<sup>10</sup> obtained a more complicated solution, with slightly different results for a junction which is closely reproduced in practice by flowing one solution down each side of a mica strip with a hole through which the solutions mix by diffusion, while constant composition is maintained at each end of the hole by the flow<sup>11</sup>. Taylor<sup>12</sup> and Guggenheim<sup>13</sup> have attempted the solution for the junction which is best approximated in practice by stopping the flow in a flowing junction apparatus. At zero time there is a plane separating the first solution from the second, but the breadth of the junction increases as the solutions diffuse into each other. However, the relatively simple treatment of Henderson is also the one of greatest practical importance.

All of these treatments agree that the breadth of the junction, that is the distance from the first solution to the second, has no effect upon the electromotive force. Yet most students of liquid junction have found that the breadth of the junction does have an important effect. An explanation of this discrepancy is obtained from the observation of Scatchard and Buehrer<sup>14</sup> that there is an effect of breadth of junction in hydrochloric acid concentration cells, in which the composition of each layer must depend upon a single variable since there are only two components. Their explanation is that the heat change produced by diffusion of acid from the concentrated solution to the dilute kept the junction at a temperature different from that of the rest of the cell. This condition is contrary to the assumption underlying equation 62. This explanation

<sup>6</sup> Clark, W. M., "The Determination of Hydrogen Ions", The Williams & Wilkins Co., Baltimore, Md., p. 296 (1928).

<sup>7</sup> Lamb, A. B., and Larson, A. T., *J. Am. Chem. Soc.*, **42**, 229 (1920).

has been confirmed by Hamer's work with sulfuric acid<sup>15</sup>. He found that the magnitude of the effect depended entirely upon the magnitude of the heat of mixing of the two solutions, which can be varied greatly for sulfuric acid. These results were obtained with the junction immersed in a large liquid thermostat capable of carrying away the heat rapidly. If the junction is surrounded by air, even though it is thermostated, the effects would probably be much bigger. The Clark type of junction now in general use is so broad that the difficulties are greatly diminished, but it is possible that an efficient liquid thermostat around the junction would greatly increase the precision of the measurements.

The measurements with concentration cells give an insight into the possibility of measuring equilibrium with respect to one process in a system which is not in equilibrium with respect to some other process. In these cells the change in state produced by the electrolytic process is the transfer of solute from one concentration to another. Exactly the same change of state is being produced irreversibly by diffusion. Yet electromotive force measurements can probably be made more accurately with these cells than with any others, and the limit on the permissible rate of the irreversible process appears to be the violation of the condition of uniform temperature rather than that of minimum energy. The behavior of these cells encourages us to apply thermodynamic treatments to other systems in which there is a slow change even though the conclusions can be derived rigidly only for systems which are at equilibrium with respect to every change.

#### Single Ion Activities, Salt Bridges and pH

It is obvious that the free energy change at a single electrode involves the chemical potential of a single ion species. So does the free energy change at a liquid boundary. The total change in any cell, however, involves only such sums of  $\sum_i \nu_i \mu_i$  as correspond to  $\sum_i \nu_i z_i = 0$ . This is most easily seen from the fact that the net charge is zero at all points in the cell. Moreover, the electrical potential difference at a single electrode or at a liquid boundary cannot be given any exact significance, but the electrical potential difference between two pieces of like metal at the extremes of the cell may be exactly determined. So neither term in equation 61 has exact meaning for an electrode or liquid boundary, but both terms do for a complete cell. The consideration of individual ion activities may be avoided by adding to equation 64 the right hand side of equation 66 and subtracting the left hand side.

The electrolytic cell most often measured in protein chemistry consists of a hydrogen electrode, or a quinhydrone or a glass electrode, in the solution to be investigated ("), with a saturated potassium chloride bridge to a standard electrode. For simplicity we will take this standard as a calomel electrode in saturated KCl (').

$M, H_2(P_{H_2}), \text{ solution } ("), KCl \text{ (sat.)}, Hg_2Cl_2(s), Hg(l).$

From equations 61, 62, 64 and 66, with the hydrogen ion chosen as the  $k'$ th in the last, we obtain after grouping terms:

$$\begin{aligned} \xi = \xi_0 + \frac{RT}{2f} \ln P_{H_2} - \frac{RT}{f} \ln m''_{H^+} + \frac{RT}{f} \int'' [\sum_i u_i^0 dm_i / \sum_j m_j u_j^0 z_j \\ + \sum_i t_i' d \ln m_i + \sum_i t_i (d \ln \gamma_i - z_i d \ln \gamma_{H^+})] \end{aligned} \quad (67)$$

$$\xi_0 = [\mu_{H^+}^0 + \mu_{Cl^-}^0 + \mu_{Hg}^0 - \frac{1}{2}\mu_{H_2}^0 - \frac{1}{2}\mu_{Hg_2Cl_2}^0] + RT \ln m'_{Cl^-} + RT \ln \gamma'_{H^+} \gamma'_{Cl^-}. \quad (68)$$

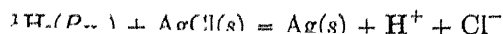
It is apparent that  $\xi_0$  depends upon the nature of the standard electrode and that it is a function of the temperature but that it does not depend upon the nature of the second solution. We may also write equation 67

$$\xi - \xi_0 - \frac{2.3 RT}{2f} \log P_{H_2} = \frac{2.3 RT}{f} [-\log m''_{H^+} + \lambda''] = \frac{2.3 RT}{f} \text{pH} \quad (69)$$

$\lambda''$  is an abbreviation for  $\frac{1}{2.3}$  times the integral of equation 67, and equation 69 is the definition proposed by Clark (6, p. 480) for pH. The value of  $\xi_0$  is determined by measuring the electromotive force with some solution (') in which the pH is fixed by definition. If  $\lambda''$  were zero, the pH by this pragmatic definition would be equal to  $-\log m''_{H^+}$ . However,  $\lambda''$  does not correspond to the integral of equation 62 or 64. It also includes  $\log \gamma'_{H^+}$ , which is included in  $\xi_0$ , and  $-\log \gamma''_{H^+}$ . So if this integral were zero, the pH would be equal to  $-\log a''_{H^+}$ .

The saturated potassium chloride bridge is avoided by physical chemists whenever possible, partly because of the uncertainties involved in  $\lambda''$ , and partly because the measurements have not been considered as precise as those with some other cells. We have already discussed the precision of the experimental measurements. There are two substitutes in common use. The first is to use the cell without liquid junction

$M, H_2, \text{ solution } ("), AgCl(s), Ag(s), \text{ for which the cell reaction is:}$



must be known, and some chloride is added if necessary. The electromotive force of this cell is given by

$$\begin{aligned}\mathcal{E} - \mathcal{E}_0 &= \frac{2.3 RT}{2\mathcal{F}} \log P_{H_2} + \frac{2.3 RT}{\mathcal{F}} \log m_{Cl^-} \\ &= \frac{2.3 RT}{\mathcal{F}} (-\log m_{H^+} - \log \gamma_{H^+} \gamma_{Cl^-})\end{aligned}\quad (70)$$

This method is used particularly for synthetic solutions in which it is useful to vary the concentration and extrapolate to zero concentration so that  $\log \gamma_{H^+} \gamma_{Cl^-}$  is zero.

The second method is the use of an indicator, which depends upon the reaction  $IH^{n'+1} = I^{n'} + H^+$

$$\log m_{H^+} = \log K - \log (m_I/m_{IH}) - \log \gamma_I \gamma_{H^+} / \gamma_{IH} \quad (71)$$

$n'$  is the valence of the basic form of the indicator, which may be any integer, positive, zero or negative. It is the ratio  $(m_I/m_{IH})$  which is determined colorimetrically. It is customary to define  $K'$  by the equations

$$\log K' = \log (m_I m_{H^+} / m_{IH}) = \log K - \log \gamma_I \gamma_{H^+} / \gamma_{IH} \quad (72)$$

If  $\log K'$  is measured at several concentrations, it is possible to extrapolate to  $\log K$  at infinite dilution. The effect of ionic strength on this difference depends, for substances which make satisfactory indicators, largely upon the value of  $n'$ . By the Debye theory, the term in  $\log \gamma_I \gamma_{H^+} / \gamma_{IH}$  proportional to the square root of the ionic strength is proportional to  $(n' + 1)^2 - n'^2 - 1 = +2n'$ , which may be positive, zero or negative. It is  $n'$  times the corresponding term in  $\log \gamma_{H^+} \gamma_{Cl^-}$ . Comparison of  $\lambda''$  in equation 69 shows that the term proportional to the square root of the ionic strength is half the term in  $\log \gamma_{H^+} \gamma_{Cl^-}$ . This is the relation which should hold if this term is the same for  $\log \gamma_{H^+}$  as for  $\log \gamma_{Cl^-}$ , and if  $\lambda''$  is equal to  $-\log \gamma_{H^+}$ .

Since the individual ion activities are undetermined, we may make one arbitrary rule about them. The rule usually made for electrode potentials is that the molal electrode potential of the hydrogen electrode is zero, or more precisely, that  $\mu_{H^+}^0 = \frac{1}{2} \mu_{H_2}^0$  at all temperatures. Several rules have been suggested for the variation of the potentials of single ion species with the composition. The simplest is that pH as defined from measurements of the electromotive force of a cell with saturated potassium chloride bridge

tivity. Then for any electrode written on the left side of a cell with a saturated potassium chloride bridge and standard electrode on the right,

$$\xi - \xi_0 = - \frac{RT}{\mathcal{F}} \sum_i \nu_i \ln a_i \quad (73)$$

in which  $\nu_i$  is the number of moles of the  $i$ 'th species produced at the left hand electrode per equivalent of positive electricity passed from left to right.  $\xi_0$  for any cell will differ from  $\xi_0$  for the cell with a hydrogen electrode by the molal potential of the electrode used in it.

### Osmotic Pressure

If there is a membrane separating two parts of a system which prevents the transfer of one or more components from one part to the other, there is no necessity that the potentials of these substances or the pressure be the same in the two parts. If there is another group of substances which can be transferred, their potentials, like the temperature, must be the same on the two sides. We consider first the case in which there are only two components, the first of which can pass through the membrane while the second cannot. If the two sides of the membrane are distinguished by primes or double primes, the conditions of equilibrium are:  $T' = T''$ ,  $\mu_1' = \mu_1''$ . We are interested in determining the relation between the potential in the second phase when at the same pressure as the first,  $(\mu_1'')'$ , and the pressure difference at equilibrium,  $P' - P''$ .

$$(\mu_1')_{P'} = (\mu_1'')_{P''} = (\mu_1'')_{P'} + \int_{P'}^{P''} \bar{V}_1'' dP \quad (74)$$

$$(\mu_1'')_{P'} - (\mu_1')_{P'} = RT \ln (a_1''/a_1')_{P'} = - \int_{P'}^{P''} \bar{V}_1'' dP \quad (75)$$

If  $\bar{V}_1''$  is a constant, this may be integrated to give

$$P'' - P' = \frac{-RT}{\bar{V}_1''} \ln (a_1''/a_1')_{P'} \quad (76)$$

If the primed solution contains none of the second component and the other solution is so dilute in this component that it may be considered ideal, then  $-\ln (1 - N_2'')$  may be replaced by  $N_2''$  and  $n_2''/(n_1'' + n_2'')\bar{V}_1''$  by  $n_2''/V'' = C_2''$ ,

$$P'' - P' = RTC_2'' \quad (77)$$

which is Van't Hoff's law for the osmotic pressure, if the osmotic pressure is defined as the pressure difference between the solution when

the solution is in equilibrium with the solvent through a membrane impermeable to the solute.

### Donnan Equilibrium

The situation is more complicated if there are more than two components, and particularly if a solution contains ions, some of which may be transferred and others may not. The chemical potential of an ion need not be the same on the two sides, but for each combination of diffusible ions such that  $\sum_i \nu_i z_i = 0$ ,  $\sum_i \nu_i \mu_i' = \sum_i \nu_i \mu_i''$  and

$$\sum_i \nu_i [\mu_i'' - \mu_i']_{P'} = RT \sum_i \nu_i \ln (a_i''/a_i')_{P'} = - \int_{P'}^{P''} \sum_i \nu_i \bar{V}_i'' dP \quad (78)$$

We will first follow the customary procedure, and assume that every activity coefficient is unity and that the integral may be considered zero for every component except the solvent. Then

$$\sum_i \nu_i \ln m_i''/m_i' = 0 \quad \text{if} \quad \sum_i \nu_i z_i = 0 \quad (79)$$

It follows that  $(m_i''/m_i')/z_i$  is the same for every species which can be transferred. If we designate by Greek subscripts those species which cannot be transferred, and if there are none of them in the solution designated by single primes, the law of electrical neutrality demands that

$$\sum_i m_i' z_i = 0$$

$$\sum_i m_i'' z_i + \sum_{\lambda} m_{\lambda}'' z_{\lambda} = 0.$$

Further progress requires a knowledge of the conditions. We will consider only the simplest possible case of three ions, of which the first two are at equilibrium through a membrane impermeable to the third, and the third is all on the side of this membrane designated by a double prime, and with  $z_1 = -z_2 = 1$ . Then

$$m_2' = m_1', \quad m_2'' = m_1'' + z_3 m_3''$$

$$(m_1')^2 = m_1''(m_1'' + z_3 m_3'')$$

$$\begin{aligned} P'' - P' &= RT [2m_1'' + (1 + z_3)m_3'' - 2\sqrt{m_1''(m_1'' + z_3 m_3'')}] \\ &= RT \left[ m_3'' + 2m_1'' \left( 1 + \frac{z_3 m_3''}{2m_1''} - \sqrt{1 + \frac{z_3 m_3''}{m_1''}} \right) \right] \end{aligned} \quad (80)$$

The limits of  $P'' - P'$  are  $RTm_3''$ , when  $z_3 m_3''/m_1''$  is very small, and  $(1 + z_3)RTm_3''$ , when  $z_3 m_3''/m_1''$  is very large. We may generalize this

fusible ions the pressure difference corresponds to the concentrations of salts of the non-diffusible ions with whatever ions of the opposite sign are present. The hydrogen and hydroxyl ions from the water distribute like other ions. Near the lower limit the distribution ratios are large so that the solution containing the non-diffusible ions may become strongly enough basic or acidic to form weak bases or acids from the non-diffusible ions. This phenomenon is known as "membrane hydrolysis."

In the study of protein solutions it is often necessary to carry out the measurements in a buffer solution, and to make corrections for the distribution of diffusible ions across the membrane. The osmotic pressure method has the advantage over the depression of the freezing point or vapor pressure that it counts only the non-diffusible ions plus a portion of the ions of opposite sign necessary to neutralize them. The other methods count every solute particle including the excess of salts of diffusible ions. It has a further advantage that the effect of a very small concentration difference is much more easily measured. In water solution the limiting ratio of  $dP/P$  to  $dM$  for the vapor pressure is 0.018, so the ratio of  $dP$  to  $dM$  varies from 0.07 at  $0^\circ$  to 14 at  $100^\circ$  if the pressure is measured in millimeters of mercury. For the osmotic pressure the ratio of  $dP$  to  $dM$  varies from 17,000 at  $0^\circ$  to 23,000 at  $100^\circ$  for pressures in millimeters of mercury. For the freezing point depression, the ratio of  $dT$  to  $dM$  is 1.86. So 1 mm. in the osmotic pressure corresponds roughly to  $0.0001^\circ\text{C}$ . A reading of a tenth of a millimeter on a water manometer corresponds to less than a millionth of a degree in the freezing point depression. The chief disadvantage of the osmotic pressure method is that there are only a limited number of solutes for which a semipermeable membrane can be made.

This unequal distribution of diffusible ions across a membrane is known as the Donnan membrane equilibrium<sup>16</sup>. The effect of activity coefficients different from unity has been discussed by Donnan and Guggenheim<sup>17</sup>. We should also consider the value of the integral of  $\bar{V}_i dP$ , which we have so far considered as zero for diffusible solutes. This assumption is justified for solutes with molal volumes the same as that of water, for the ratio  $a'_i/a''_i$  is increased only 1.8% per mole of undiffusible solute in a kilogram of water. However, if the molal volume is a thousand times that of water, which corresponds roughly to a protein of molecular weight 30,000, the activity ratio is increased 1.8% per 0.001 mole per kilogram. The corresponding pressure difference is only twenty millimeters of mercury.

### Membrane Potentials

If we set up a cell with two electrodes of the same kind, one on either



on the chemical potentials of the materials which react at the electrode but are not in the solution, even though the concentration of the ion which reacts at the electrode may be very different because of the presence on one side of the membrane of an ion to which the membrane is impermeable, and, except for gas electrodes, the pressure difference is negligibly small. That the potential is zero may be seen from the fact that the cell process must correspond to a transfer such that  $\sum_i \nu_i z_i = 0$ , and for such transfers  $\sum_i \nu_i \mu_i' = \sum_i \nu_i \mu_i''$ . With the same arbitrary assumptions we made to determine single ion potentials, we may determine the electrical potential difference across the membrane as the electromotive force of a cell in which two identical electrodes are connected to the solution on the two sides of the membrane through saturated potassium chloride bridges. This electromotive force is  $(RT/z_k) \ln (a_k'/a_k'')$  for any ion which is at equilibrium on the two sides.

#### Gravitational and Centrifugal Fields

The simplest treatment of equilibrium in a gravitational field is to omit the electrical and surface terms in equation 61 and to take

$$dW = \sum_i \bar{W}_i dn_i \quad (81)$$

in which  $\bar{W}_i$  is the molecular weight of the  $i$ 'th component. Then equation 62 becomes

$$\sum_i (\mu_i + G\bar{W}_i) dn_i = 0 \quad (82)$$

Since a possible variation is the transfer of  $n_k$  moles of the  $k$ 'th component through a change in the gravitational potential, it follows that at equilibrium for each component

$$\mu_k + G\bar{W}_k = \text{constant} \quad (83)$$

or

$$d\mu_k/dG = -\bar{W}_k \quad (84)$$

We assume that the chemical potentials are functions of the temperature, pressure and composition, so that, at equilibrium and therefore at constant temperature,

$$\begin{aligned} -\bar{W}_k &= d\mu_k/dG = (\partial\mu_k/\partial P)_{T, n} dP/dG + \sum_j (\partial\mu_k/\partial N_j)_{T, P} dN_j/dG \\ &= \bar{V}_k dP/dG + \sum_j (\partial\mu_k/\partial N_j)_{T, P} dN_j/dG \end{aligned} \quad (85)$$

Summing over all the components yields

$$-(\sum_k n_k \bar{W}_k) = \sum_k n_k \bar{V}_k dP/dG + \sum_k \sum_j (\partial\mu_k/\partial N_j)_{T, P} dN_j/dG$$

By equation 1, the second term in equation 86 is zero, so

$$dP/dG = -W/V = -\rho \quad (87)$$

if  $\rho$  is the density. Inserting this value in equation 85 yields

$$\sum_j (\partial \mu_k / \partial N_j) dN_j / dG = -(\bar{W}_k - \bar{V}_{k\rho}) \quad (88)$$

in which  $-(\bar{W}_k - \bar{V}_{k\rho})$  is the apparent mass of a mole of the  $k$ 'th component in the system. In an ideal solution this reduces to

$$d \ln N_k / dG = -(\bar{W}_k - \bar{V}_{k\rho}) / RT \quad (89)$$

in a non-ideal solution we may write

$$\frac{d \ln m_k}{dG} + \sum_j \left( \frac{\partial \ln \gamma_k}{\partial m_j} \right)_{T, P, m} \frac{d m_j}{dG} = -(\bar{W}_k - \bar{V}_{k\rho}) / RT \quad (90)$$

in which the solvent is omitted from the summation.

The application of these equations to the ultracentrifuge, in which  $G$  is taken as the centrifugal potential, will be considered in chapter 19. Here we are interested in the problem of whether the earth's gravitational field may be neglected in the study of protein solutions. For an approximate solution we take equation 89, assume that for a protein  $\bar{V}_k = 0.75 \bar{W}_k$  and that  $\rho = 1$ . Then

$$d(\ln N_k) / dh = -(0.25 \times 980 / 8.3 \times 10^7 \times 300) \bar{W}_k = \bar{W}_k \times 10^{-8} \text{ cm}^{-1}$$

If the molecular weight is less than a hundred thousand, the change in concentration at equilibrium is less than 0.1% per centimeter, but if the molecular weight is greater than ten million, the change is more than ten per cent per centimeter. So for large proteins the effect may be very appreciable within the dimensions of an ordinary apparatus.

### Surface Tension

In considering surface energy we will assume that there is but a single kind of surface, and that the surface tension is a function of the temperature, pressure and compositions of the phases which meet at the surface (or surface concentration of a substance not soluble in either phase). The extension to several kinds of surfaces is obvious, but the second assumption excludes from our treatment the consideration of the effect of curvature on surface tension. It has the advantage that it makes the pressure, like the temperature, uniform throughout a system at equilibrium. Gibbs's treatment of surfaces employs a much more specific physical picture than the rest of his thermodynamics, and it would be too long for discussion

treatment includes the effect of curvature, we will define a surface concentration somewhat different from his. For a two component system both reduce to the same thing, and it is possible that our definition is not less useful than that of Gibbs for more complicated systems.

Since the pressure and the temperature are uniform throughout, it is convenient to treat the free energy, the differential equation for which becomes

$$dF = -SdT + VdP + \sum_i \mu_i dn_i + \sigma d\mathcal{A} \quad (91)$$

We start with the fact that

$$(\partial\sigma/\partial n_k)_{T, P, n, \alpha} = (\partial^2 F / \partial \mathcal{A} \partial n_k) = (\partial^2 F / \partial n_k \partial \mathcal{A}) = (\partial \mu_k / \partial \mathcal{A})_{T, P, n}$$

We then use successively the cyclical rules for partial derivatives with rotation of subscripts and with constant subscripts and invert to obtain

$$\begin{aligned} (\partial\sigma/\partial n_k)_{T, P, n, \alpha} &= -(\partial n_k / \partial \mathcal{A})_{T, P, n, \mu_k} (\partial \mu_k / \partial n_k)_{T, P, n, \alpha} \\ (\partial\sigma/\partial \mu_k)_{T, P, n, \mu_k, \alpha} &= -(\partial n_k / \partial \mathcal{A})_{T, P, n, \mu_k} = -\Gamma_k \end{aligned} \quad (92)$$

We define the surface concentration  $\Gamma_k$  by equation 92. It is the amount of the  $k$ 'th component which must be added per unit increase of surface to keep its chemical potential constant when the quantities of the other components are kept constant. For a two component system, if  $\mu_1$ ,  $T$  and  $P$  are constant in a bulk phase,  $N_1$  must be also, and  $\Gamma_2$  may be written  $(\partial n_2 / \partial \mathcal{A})_{T, P, N}$ . This definition shows that the composition of a bulk phase is kept constant as the surface is increased. So  $\Gamma_k$  may well be called the surface concentration, although there is no necessity to say where the material is situated with respect to any geometrical surface. For more complex systems, the concentration of other substances in the bulk phases will also change with the extent of the surface, and the relation is not so simple.

Gibbs's relation is

$$-\Gamma_{k(j)} = (\partial\sigma/\partial \mu_k)_{T, P, n_j, \mu_j \neq \mu_i \text{ or } \mu_k} \quad (93)$$

in which  $\Gamma_{k(j)}$  is the difference in quantity of the  $k$ 'th component per unit surface in the real system from that in an imaginary system which has the composition of one bulk phase on one side of a surface and that of the other bulk phase on the other side of the surface. This surface is parallel to the real interface and drawn so that the difference in quantity of the  $j$ 'th component is zero.

In either case we may replace the chemical potential by the activity to

which in dilute solutions becomes

$$-\Gamma_k = (C_k/RT)\partial\sigma/\partial C_k \quad (95)$$

The most important qualitative relation is that in dilute solutions  $\Gamma_k$  cannot have a large negative value, so the surface tension cannot increase rapidly with the concentration in dilute solutions. It may, however, have a large positive value, so the rate of decrease of the surface tension may be much greater. If the surface concentration is zero, the surface tension-concentration curve passes through a minimum or a maximum. From the first relation we see that a minimum is much more probable than a maximum.

### Solubility and Particle Size<sup>3</sup>

The variation of solubility with particle size depends upon the variation of the surface between two phases with change in the size of particles of the dispersed phase. If  $\mu_k$  is the chemical potential in the dispersed phase and  $\mu'_k$  that in the continuous phase,

$$\mu'_k = \mu_k + \sigma(\partial\mathcal{A}/\partial n_k)_{T, P} \quad (96)$$

or

$$\ln a'_k = \mu_k/RT + (\sigma/RT)(\partial\mathcal{A}/\partial n_k)_{T, P} \quad (97)$$

If the shape of the particles remains constant, we may write  $d\mathcal{A}$  as  $4dV/\delta$ , in which  $\delta$  is a measure of the particle size. For a sphere,  $\delta$  is the diameter, for a cube it is the cube edge; and for a rectangular prism with edges  $a$ ,  $b$  and  $c$ ,  $3/\delta = 1/a + 1/b + 1/c$ . Then

$$\ln a'_k = \ln a_k^0 + 4\sigma\bar{V}_k/RT\delta \quad (98)$$

$\ln a_k^0 = \mu_k/RT$  is obviously the value of  $\ln a'_k$  for an extremely large particle. For water drops, with  $\sigma = 80$  ergs/cm<sup>2</sup> and  $\bar{V}_k = 18$  cm<sup>3</sup>,  $\log a'_k = \log a_k^0 + 1 \times 10^{-7}\delta$ . The effect is small until  $\delta$  is less than a millionth of a centimeter, but the pressure is ten times as great from a drop a ten-millionth of a centimeter in diameter as from an extremely large drop. The same equation applies to the activity in a capillary if  $-\delta$  be taken as the diameter of a circular capillary or  $-\delta/2$  be taken as the mean radius of curvature of the meniscus in a pore of any shape.

The effect of particle size on the solubilities of salts in water indicates that the surface tension at the salt-solution interface is of the order of a thousand ergs per square centimeter.<sup>18</sup> The surface tension at protein-solution interfaces is not known, but the protein molecules are so large

above which the effect becomes negligible, is to be smaller than the size of particles usually used for solubility measurements. The effect can be nullified only partially by keeping the particle sizes always the same in a series of measurements of solubilities in different media. If we designate the second solution by double primes, and distinguish the surface tensions in the same way,

$$\ln (a_k''/a_k') = (4\bar{V}_k/RT)(\sigma''/\delta'' - \sigma'/\delta') \quad (99)$$

Even if the  $\delta$ 's are the same, there will be a difference in the activities if the surface tensions are different.

At complete equilibrium the dispersed phase would be a single particle. We usually trust the system to reach a state in which the size of the dispersed particles is so large that the difference in energy from this equilibrium state is negligibly small. It is excellent practice, however, to check the time necessary to justify this trust. The customary assumption that the properties of a system depend only on the temperature, pressure, composition and state of aggregation depends upon the trust that the system spontaneously reaches equilibrium with respect to many other variables which would otherwise affect the properties. For solids there are, in addition to the particle size, the stable crystalline lattice, the stable shape, which may be related to different surface tensions on different crystal faces, and freedom from strain. In pure liquids there is orientation relative to phase boundaries, or merely to adjacent parts of the same phase. Equilibrium may be attained very slowly if there is a favored orientation as in liquid crystals. In solutions there are also composition differences at surfaces or in a potential field. Although thermodynamics is generally applicable, the equations we use to apply it are valid only under certain restrictions, and we must check carefully in each new case to determine whether these restrictions correspond to the conditions of our experiments.

#### SIMPLE ELECTROSTATIC THEORY

The applications of the theories of Debye<sup>19, 20</sup> concerning solutions and of the extensions of these theories to solutions of dipolar ions constitute such a large part of the chapters which follow that it is expedient to derive here the equations for the simplest possible model and to discuss in a later chapter the more complicated models. We assume that each ion is a rigid sphere of radius  $b$ , within which the distribution of the ionic charge,  $\epsilon z_i$ , and the dielectric constant are spherically symmetrical, independent of the medium and not distorted by an electrical field; and outside of which the

cannot approach closer than a distance  $a_{ij}$ , which is not smaller than  $b_i$  or  $b_j$ , but at greater distances the interaction is given completely by the electrostatic forces between the net ionic charges. We shall assume that the distance  $a_{ij}$  is the same for any pair of ions, and shall therefore drop the subscripts. Some of these assumptions will be made more general later.

The solution as a whole is neutral. Therefore the location of a charge  $\epsilon z_i$  within the sphere of radius  $b_i$  requires that there be another charge  $-\epsilon z_i$  outside this sphere. Our first task is to consider the average distribution of this charge, which Debye calls the "ion atmosphere." It cannot be uniformly distributed. We may determine the concentrations of ions in an electrical field just as we did those of masses in a gravitational field to obtain equation 84 and those which follow. Omitting gravitational and surface terms from equation 59, substituting the symbol  $\psi$  for  $\xi$  to correspond to later chapters, and utilizing the fact that  $de = \sum_i \epsilon_i z_i dn_i$ , in which  $\epsilon$  is the charge on a proton,  $z_i$  is the valence of the  $i$ 'th species and  $n_i$  is the number of molecules of that species ( $n'_i = n_i/N$ , in which  $N$  is Avogadro's number), we obtain from equation 60

$$\sum_i (\mu_i/N + \psi \epsilon z_i) dn'_i = 0 \quad (100)$$

Since the transfer of  $dn'_k$  molecules of the  $k$ 'th species is a possible variation

$$kT d \ln a_k/d\psi = d(\mu_k/N)/d\psi = -\epsilon z_k \quad (101)$$

in which  $k = R/N$  is Boltzmann's constant. So

$$a_k = a_k^0 e^{-\psi \epsilon z_k/kT} \quad (102)$$

in which  $a_k^0$  is the activity of the  $k$ 'th component at a point where  $\psi$  is zero. Substituting concentrations for activities in the dilute solutions, we obtain

$$c_k = c_k^0 e^{-\psi \epsilon z_k/kT} \quad (103)$$

which is Boltzmann's distribution law for an electrical field.

Poisson's equation may be written

$$\nabla^2 \psi = -4\pi\rho/D \quad (104)$$

in which  $\rho$  is the electrical density,  $D$  is the dielectric constant, and  $\nabla^2$  is the Laplacian operator.

$\nabla^2 \psi = \partial^2 \psi / \partial x^2 + \partial^2 \psi / \partial y^2 + \partial^2 \psi / \partial z^2$  in Cartesian coördinates

$$= \frac{1}{r^2} \frac{\partial}{\partial r} \left( r^2 \frac{\partial \psi}{\partial r} \right) + \frac{1}{r^2 \sin^2 \theta} \frac{\partial}{\partial \phi} \left( \sin^2 \theta \frac{\partial \psi}{\partial \phi} \right) + \frac{1}{r^2 \sin^2 \theta} \frac{\partial^2 \psi}{\partial \theta^2} \quad (105)$$

Combining equations 103 and 104, and letting  $c_i$  and  $c_i^0$  be concentrations in molecules per cubic centimeter

$$\nabla^2 \psi = -(4\pi\epsilon/D)\Sigma_i c_i z_i = -(4\pi\epsilon/D)\Sigma_i c_i^0 z_i e^{-\psi u_i/kT} \quad (106)$$

This equation cannot be integrated analytically. So we expand the exponential and retain only the first two terms, to give

$$\nabla^2 \psi = -(4\pi\epsilon/D)\Sigma_i c_i^0 z_i + (4\pi\epsilon^2/DkT)(\Sigma_i c_i^0 z_i^2) \psi = \kappa^2 \psi \quad (107)$$

The first term is zero because the net charge in the solution,  $\Sigma_i c_i^0 z_i \epsilon$ , is zero. The quantity  $\kappa$  is defined by this equation as

$$\begin{aligned} \kappa &= \sqrt{(4\pi\epsilon^2/DkT)\Sigma_i c_i^0 z_i^2} = \sqrt{(4\pi N\epsilon^2/1000 DkT)\Sigma_i C_i^0 z_i^2} \\ &= \sqrt{(4\pi N\epsilon^2/1000 DkT)\Gamma} \end{aligned} \quad (108)$$

$\Gamma = \Sigma_i C_i^0 z_i^2$  is the ionic concentration.  $C_i$  is the concentration of the  $i$ 'th species in moles per liter. Except that the concentration units are moles per liter rather than moles per kilogram of solvent,  $\Gamma$  is twice the ionic strength,  $\Sigma_i m_i z_i^2/2$ .  $\kappa$  is a very important quantity in the Debye theory for reasons which will appear later.

### The Debye-Hückel Equation

It is convenient to use the spherical coordinate form of the Laplacian and our assumption of spherical symmetry makes  $\partial\psi/\partial\theta$  and  $\partial\psi/\partial\phi$  both zero, so

$$\nabla^2 \psi = \frac{1}{r^2} \frac{d}{dr} \left( r^2 \frac{d\psi}{dr} \right) = (1/r)^4 d^2 \psi / d(1/r)^2 = \kappa^2 \psi \quad (109)$$

The solution, which is most easily verified by differentiation, is

$$\psi = \frac{Ce^{-\kappa r}}{r} + \frac{C'e^{\kappa r}}{r} \quad (110)$$

The two parameters,  $C$  and  $C'$ , are determined from the boundary conditions. Since  $\psi$  approaches zero and  $e^{\kappa r}/r$  approaches infinity as  $r$  approaches infinity,  $C'$  must be zero. At the boundary  $r = a$ ,  $\psi$  and  $d\psi/dr$  must each be continuous. When  $r \geq a$

$$\psi = Ce^{-\kappa r}/r \quad (111)$$

$$d\psi/dr = -Ce^{-\kappa r}(1 + \kappa r)/r^2 \quad (112)$$

and when  $b \leq r \leq a$ , Coulomb's law may be used with the central charge, since there are no charges within this shell, to give

From equations 112 and 113 when  $r = a$

$$-Ce^{-\kappa a}(1 + \kappa a)/a^2 = -\epsilon z_i/Da^2 \quad (115)$$

$$C = \epsilon z_i e^{\kappa a}/D(1 + \kappa a) \quad (116)$$

and when  $r \geq a$

$$\psi = \epsilon z_i e^{-\kappa(r-a)}/D(1 + \kappa a)r \quad (117)$$

So, when  $r = a$ , using also equation 114,

$$\psi_a = \epsilon z_i/Da(1 + \kappa a) = (\epsilon z_i/D)[1/a - \kappa/(1 + \kappa a)] = \epsilon z_i/Da + C'' \quad (118)$$

and, therefore,

$$C'' = -\epsilon z_i/D(1 + \kappa a) \quad (119)$$

and from equations 114 and 119, when  $r = b$

$$\psi_b = (\epsilon z_i/D)[1/b_i - \kappa/(1 + \kappa a)] \quad (120)$$

In this simple case we may also use the method of MacInnes<sup>21</sup> and determine  $C$  from the fact that the total charge outside the sphere  $r = a$  is  $-\epsilon z_i$ . Combining equations 104, 107, and 111 gives

$$\rho = (-D/4\pi)\kappa^2\psi = (-D/4\pi)Ce^{-\kappa r}/r \quad (121)$$

$$-\epsilon z_i = \int_a^\infty 4\pi r^2 \rho \, dr = \int_a^\infty -DCe^{-\kappa r} r \, dr = -DCe^{-\kappa a}(1 + \kappa a) \quad (122)$$

This yields the same value of  $C$  as the method used above. The first method is more closely related to those which must be used in the more complicated cases.

The net charge per unit thickness in a spherical shell is

$$4\pi r^2 \rho = -[\epsilon z_i e^{\kappa a}/(1 + \kappa a)]e^{-\kappa r}r \quad (123)$$

This quantity passes through a maximum when  $r = 1/\kappa$ . Perhaps the fact that more of the ionic atmosphere is at a distance  $1/\kappa$  than at any other distance is the best justification for calling  $1/\kappa$  the "thickness of the ionic atmosphere". Three-quarters of the charge is at a distance greater than  $1/\kappa$  at the limit  $\kappa a = 0$ , and this fraction increases to unity at the limit  $\kappa a = 1$ .

It is evident from equation 120 that  $\epsilon z_i/Db_i$  is the potential on the surface of the sphere  $b_i$  due to the ion itself, and that  $-\epsilon z_i\kappa/D(1 + \kappa a)$  is the potential due to the ion atmosphere. For the calculation of conductance, transference or electrophoresis and other quantities which depend



upon the transfer of the ions, the potential as given in equation 118 or 120 is used directly. For the calculation of equilibrium relations, we imagine that there is a process by which the charge may be added to, or removed from, the ion reversibly. Then  $W_{ei}$ , the work of bringing this charge to the surface of the sphere of radius  $b_i$ , is

$$W_{ei} = \int_0^{ez_i} \psi_{bi} de \quad (124)$$

The work of changing the distribution of the charge within the sphere is independent of the conditions if the restrictions given above are maintained. We consider first the change of state from one in which all the ions are at infinite dilution in the standard solvent to one in which they have the given concentrations in a given solvent. Our process will be to remove the charges of the ions in the standard solution, to transfer the discharged ions to the proper amount of the given solvent, and then to add the charges again. The discharging and charging processes are so carried out that at any instant each ion has the same fraction,  $\lambda$ , of its final charge. The number of ions of the  $i$ 'th species in the system is  $n'_i$ .

In the first step the electrical work is

$$\begin{aligned} -W_e^0 &= -\sum_i n'_i W_{ei}^0 = -\sum_i n'_i \int_0^{ez_i} \psi_{bi} de = -\sum_i n'_i \int_0^{ez_i} (e/D_o b_i) de \\ &= -\sum_i n'_i (\epsilon^2 z_i^2 / D_o b_i) \int_0^1 \lambda d\lambda = -\sum_i n'_i \epsilon^2 z_i^2 / 2 D_o b_i \end{aligned} \quad (125)$$

There is no electrical work in the second step. In the third step

$$\begin{aligned} W_e &= \sum_i n'_i W_{ei} = \sum_i n'_i \int_0^{ez_i} \psi_{bi} de \\ &= \sum_i n'_i \int_0^1 (\epsilon^2 z_i^2 / D) [1/b_i - \lambda \kappa / (1 + \lambda \kappa a)] \lambda d\lambda \\ &= \sum_i n'_i (\epsilon^2 z_i^2 / 2D) \{1/b_i - [\kappa^2 a^2 - 2\kappa a + 2 \ln(1 + \kappa a)] / \kappa^2 a^3\} \end{aligned} \quad (126)$$

The total electrical work is

$$\begin{aligned} W_e - W_e^0 &= \sum_i n'_i \{ (\epsilon^2 z_i^2 / 2b_i) (1/D - 1/D_o) \\ &\quad - (\epsilon^2 z_i^2 / 2D) [\kappa^2 a^2 - 2\kappa a + 2 \ln(1 + \kappa a)] / \kappa^2 a^3 \} \end{aligned} \quad (127)$$

In addition to the electrical work, there will be an ideal term in the second step, and there may be non-ideal terms in any of the three steps, which we will designate as  $\delta_1$ ,  $\delta_2$  and  $\delta_3$ . Then for the whole process and for the

For the present we will assume that  $(\delta_1 + \delta_2 + \delta_3)$  is zero, which is probably much more probable than the assumption that each is zero. Later we will assume that  $(\delta_1 + \delta_3)$  is zero, but that  $\delta_2$  can be determined from the behavior of similar non-electrolytes.

In obtaining the chemical potentials by differentiation, we note that the addition of a component may change the volume and the dielectric constant, and that

$$\frac{\partial \kappa}{\partial n_k} = \frac{\kappa}{2} \left( \frac{z_k^2}{\sum_j n_j z_j^2} - \frac{\partial \ln V}{\partial n_k} - \frac{\partial \ln D}{\partial n_k} \right) \quad (129)$$

$$\begin{aligned} \mu_{ke} = \frac{\partial W_e}{\partial n_k} = \frac{N\epsilon^2}{2D} \left\{ z_k^2 \left[ \frac{1}{b_k} - \frac{\kappa}{1 + \kappa a} \right] \right. \\ \left. + Y \frac{\partial V}{\partial n_k} \sum_j C_j z_j^2 - \frac{\partial \ln D}{\partial n_k} \sum_j n_j z_j^2 \left( \frac{1}{b_j} - \frac{\kappa}{1 + \kappa a} \right) \right\} \end{aligned} \quad (130)$$

in which

$$Y = [1 + \kappa a - 1/(1 + \kappa a) - 2 \ln (1 + \kappa a)]/\kappa^2 a^3 \quad (131)$$

The first term of equation 130 depends upon the valence of the  $k$ 'th component, and is zero for non-electrolytes. It is the important term for ions, and as written corresponds to an individual ion chemical potential so that it has only formal significance; the second term depends upon the molal volume of the  $k$ 'th component and in dilute solution is important only for the solvent; the third term depends upon the effect on the dielectric constant and is the important one for non-electrolyte solutes. For an electrolyte, we will let  $\nu_+$  be the number of cations of valences  $z_+$  and  $\nu_-$  be the number of anions of valence  $z_-$  from a molecule of a salt, and let  $\nu = \nu_+ + \nu_-$ . Since the salt is electrically neutral

$$\nu_+ z_+ = -\nu_- z_- \quad (132)$$

so

$$\nu_+ z_+^2 + \nu_- z_-^2 = -\nu z_+ z_- \quad (133)$$

Similarly we define  $b_s$  by the equation

$$-\nu z_+ z_- / b_s = \nu_+ z_+^2 / b_+ + \nu_- z_-^2 / b_- \quad (134)$$

Then since  $\mu_{se} = \nu_+ \mu_{+e} + \nu_- \mu_{-e}$

$$\mu_{se} = \frac{N\epsilon^2}{2D} \left\{ -\nu z_+ z_- \left( \frac{1}{b_s} - \frac{\kappa}{1 + \kappa a} \right) \right. \quad (135)$$

The electrical contribution to the mean activity of the ions, which we will represent by  $a_{\pm e}$  is related to  $\mu_{se}$  by the equation

$$\ln a_{\pm e} = (\mu_{se} - \mu_{se}^0)/\nu RT \quad (136)$$

For a non-electrolyte solute or the solvent, or formally for a single ion species

$$\ln a_{ke} = (\mu_{ke} - \mu_{ke}^0)/RT \quad (137)$$

The activities also contain the ideal parts,  $\ln N_k/N_k^0$ , etc., and any non-ideal parts which come from  $(\delta_1 + \delta_2 + \delta_3)$ . Equations 130 and 135 contain the whole of the Debye equilibrium theory for spherical ions and for the approximation of equation 107.

The Debye-Hückel equation for the effect of electrolytes on electrolytes is obtained by letting  $D = D_0$ . We will designate by  $\kappa_0$  and  $Y_0$  the values of  $\kappa$  and of  $Y$  obtained in this manner. Then

$$\mu_{se} - \mu_{se}^0 = \frac{N\epsilon^2 \nu z_+ z_-}{2D_0} \frac{\kappa_0}{1 + \kappa_0 a} + \frac{N\epsilon^2 Y_0}{2D_0} \frac{\partial V}{\partial n_s} \sum_j C_j z_j^2 \quad (138)$$

$$\mu_{oe} - \mu_{oe}^0 = \frac{N\epsilon^2 Y_0}{2D_0} \frac{\partial V}{\partial n_0} \sum_j C_j z_j^2 \quad (139)$$

It is customary to drop the second term of equation 138. If the volumes are additive it is less than one-third the first term multiplied by the fraction of the volume occupied by the salt. If equations 138 and 139 are to be kept thermodynamically consistent, however, this term must be retained.

Hückel<sup>22</sup> has shown that for a single salt the difference between equation 135 and the first term of equation 138 may be expressed as a linear function of the concentration,

$$\mu_{se} - \mu_{se}^0 = \frac{N\epsilon^2 \nu z_+ z_-}{2D_0} \left( \frac{\kappa_0}{1 + \kappa_0 a} - \frac{dD}{D_0 b_s dC_s} C_s \right) \quad (140)$$

with  $dD/dC_s$  a constant.

If the dielectric constant is independent of the salt concentration,  $\kappa$  is proportional to the square root of the salt concentration in moles per liter of solution; but if the dielectric constant is proportional to the amount of solvent in unit volume of solution, which is a much better approximation for the behavior of aqueous solutions of many non-electrolytes,  $\kappa$  is proportional to the square root of the salt concentration in moles per liter of solvent, and therefore in moles per kilogram of solvent. So we may represent the effect of purely electrostatic action by

$$\mu_{ke} - \mu_{ke}^0 = - \frac{N\epsilon^2 z_k^2}{2D_0} \frac{\kappa}{1 + \kappa a} \quad (141)$$

We may expand equation 141 in a power series to give

$$\mu_{ke} - \mu_{ke}^0 = - \frac{N\epsilon^2 z_k^2}{2D_0} (\kappa - \kappa^2 a + \kappa^3 a^2 - \kappa^4 a^3 + \dots) \quad (143)$$

$$\ln \gamma_k = - \frac{N\epsilon^2 z_k^2}{2D_0 R T} (\kappa - \dots) = \left[ \frac{\pi N\epsilon^6 \rho_0}{1000 (D_0 k T)^3} \right]^{1/2} z_k^2 (\sqrt{I} - \dots) \quad (144)$$

Comparison of equations 31 and 144 shows that  $A$  of equation 31 is given by

$$A = \frac{2}{3} \left[ \frac{\pi N\epsilon^6 \rho_0}{1000 (D_0 k T)^3} \right]^{1/2} \quad (145)$$

It is evident that the series in 143 does not converge rapidly unless  $\kappa a$  is much smaller than unity. It is therefore often convenient to retain the denominator of equation 141, and to replace 31 and 32 by

$$\begin{aligned} \ln \gamma_k = & \frac{3}{2} A z_k^2 \sqrt{I} / (1 + a' \sqrt{I}) + 2 \sum_i B_{ik} m_i \\ & + \left( 2 \sum_i C_{ik} m_i + \frac{z_k^2}{2I} \sum_{ij} C_{ij} m_i m_j \right) \sqrt{I} + \dots \end{aligned} \quad (146)$$

$$\varphi = 1 + \frac{A}{2} Y \sum_i z_i^2 N_i + M \sum_{ij} N_i N_j \left( B_{ij} + \frac{3}{2} C_{ij} \sqrt{I} \right) + \dots \quad (147)$$

$$Y = [1 + a' \sqrt{I} - 1/(1 + a' \sqrt{I}) - 2 \ln(1 + a' \sqrt{I})]/a'^3 I \quad (148)$$

At the limit  $a' \sqrt{I} = 0$ ,  $Y$  approaches  $\sqrt{I}/3$ . For moderately small values of  $a' \sqrt{I}$ , however, the computation of  $Y$  is inconvenient, but it may easily be obtained from a table of  $a' Y$  as a function of  $a' \sqrt{I}/(1 + a' \sqrt{I})$ .<sup>23</sup> We see that  $\ln \gamma_k$  may be expressed as  $1/DT$  times a function of  $\kappa$ , which is  $1/DT$  times a function of the square root of  $1/DT$  times the ional concentration  $\Gamma$ . This is true not only for the simple spherical model but also for any model which uses only two terms in the expanded exponential of equation 107, though not of the added terms which are discussed later under Electrostatic Association. Therefore if  $(DT/D^0 T^0) \ln \gamma$  is plotted against  $(D^0 T^0/DT) \Gamma/2$  or its square root, the electrostatic terms should be independent of both dielectric constant and temperature.  $D^0$  and  $T^0$  are the dielectric constant and temperature of the standard state, usually taken as water at 25°C.

### Aqueous Solutions

There is a great advantage for calculating heats of dilution etc. in

The latest values of the universal constants, the Critical Table densities of water, and Wyman's measurements of the dielectric constants of water<sup>24</sup> are fitted by

$$A = 0.5305(1 + 1.5471 \times 10^{-3}t + 3.569 \times 10^{-6}t^2 + 2.389 \times 10^{-9}t^3) \quad \text{for water} \quad (149)$$

in which  $t$  is the centigrade temperature. Similarly

$$10^{-8}\kappa/\sqrt{I} = 0.2292(1 + 5.217 \times 10^{-4}t - 9.16 \times 10^{-6}t^2 + 8.88 \times 10^{-9}t^3) \quad \text{for water} \quad (150)$$

At 0° the first terms in  $\ln \gamma_k/z_k^2$  and in  $\log \gamma_k/z_k^2$  are  $0.7958\sqrt{I}$  and  $0.3456\sqrt{I}$ , or  $1.1254\sqrt{I/2}$  and  $0.4887\sqrt{I/2}$ .  $10^{-8}\kappa/\sqrt{I/2}$  is 0.3241.

For approximate computations it is simple to remember that in aqueous solutions the limiting slope of  $\log \gamma$  versus  $\sqrt{I/2}$  is very close to one-half, and that  $\kappa a$  is very close to one-third of  $\sqrt{I/2}$  multiplied by  $a$  in Ångström units. For many simple electrolytes  $a$  appears to lie between three and six Ångströms, so that  $\kappa a$  lies between  $\sqrt{I/2}$  and  $2\sqrt{I/2}$ . One of the practical questions arising in the study of protein solutions is the standardization of pH measurements. A simple method<sup>25</sup> is to take hundredth molal hydrochloric acid as the standard. If the activity of the hydrogen ion were given by the first term of equation 146 and equation 149, with  $\kappa a = 1.1\sqrt{I/2}$ , the pH would be 2.044 at 0°, 2.046 at 25° and 2.048 at 50°C. Unless there are specific effects which vary rapidly with the temperature, it is well within the limit of error of our other assumptions regarding pH to assume that the pH of hundredth molal hydrochloric acid is independent of the temperature<sup>26</sup>.

### Heat and Volume Changes

The model upon which we have based the Debye theory would correspond to sizes  $a$  which are rigorously constant regardless of the pressure or temperature. There is a very great advantage, however, in remembering that this is, after all, only an approximate picture of the electrolyte solution, and in keeping constant  $a' = \kappa a/\sqrt{I}$  rather than  $a$  itself. This demands that  $a$  decrease with increasing temperature or pressure, which should be expected for non-rigid ions. The compressibility assumed is one and a half times as great as that of the solvent, and for water the decrease in  $a$  from 0° to 100°C. is only five per cent. Any deviations of this assumption from the truth will appear as additional temperature variations of the parameters of the succeeding terms of equations 146 and 147. The ad

vantage of this treatment is that the only function of the temperature or pressure in the first term of equation 146 is  $A$  itself. So the partial molal heat content, heat capacity, volume compressibility, thermal expansion all have the form.

$$\bar{G}_{ke} = \bar{G}_{ke}^0 + g_e z_k^2 \sqrt{I} / (1 + a' \sqrt{I}) \quad (151)$$

The corresponding apparent molal quantities have the form

$$(\bar{G}_{ke})_{app.} = \bar{G}_{ke}^0 + \Phi_{\bar{G}_{ke}} = \bar{G}_{ke}^0 + g_e z_k^2 [\sqrt{I} / (1 + a' \sqrt{I}) - Y] \quad (152)$$

and the partial molal quantities for the solvent have the form

$$\bar{G}_{0e} = \bar{G}_{0e}^0 - w_0 \nu m g_e z_k^2 Y \quad (153)$$

For the molal heat content equations 149 and 43 yield

$$g_e = 258 \sqrt{0.5} \left[ 1 + 1.194 \left( \frac{t}{100} \right) + 0.935 \left( \frac{t}{100} \right)^2 + 0.401 \left( \frac{t}{100} \right)^3 + 0.062 \left( \frac{t}{100} \right)^4 \right]$$

if the units of  $g_e / \sqrt{I}$  are calories per mole of ion and 1 calorie = 4.186 joules. So the limiting law for the total heat of dilution per mole of salt at 0°C. is

$$\begin{aligned} \Phi_{H_2} - \Phi_{H_2}^0 &= -(2g_e/3) \nu z_+ z_- \sqrt{I} = -172 \nu z_+ z_- \sqrt{I/2} \\ &= -172 \nu z_+ z_- \sqrt{I/2} \end{aligned} \quad (154)$$

The differentiation to give the corresponding terms for the heat capacity, etc., is obvious.

The change of dielectric constant with pressure has not been measured accurately enough to give certain values for the volume change in dilution. However, if we assume that at constant temperature but varying pressure the dielectric constant is proportional to the concentration of water we obtain for aqueous solutions at 25°C.

$$\Phi_{v2} - \Phi_{v2}^0 = -0.86 \nu z_+ z_- \sqrt{I/2} \quad (155)$$

From the study of the most accurate density measurements Redlich<sup>27</sup> obtains

$$\Phi_{v2} - \Phi_{v2}^0 = -0.93 \nu z_+ z_- \sqrt{I/2} \quad (156)$$

The electrical contribution to  $\Phi_{v2}^0$  is given by differentiating  $\mu_{ke}$  at zero electrolyte concentration with respect to the pressure. From equation 130 we obtain

$$\left( \frac{d\mu_{ke}}{dp} \right) = \frac{d}{dP} N \epsilon^2 z_k^2 = \frac{-N \epsilon^2 z_k^2}{\epsilon^3} \left( \frac{d \ln D}{dP} + \frac{d \ln b_k}{dP} \right) \quad (157)$$

If we assume that  $d \ln b_k/dP = 0$  and that  $d \ln D/dP = \beta$  as for equation 155, and measure  $b_k$  in Ångströms and  $d\mu_{ks}/dP$  in cc. per mole of ion

$$\frac{d\mu_{ks}}{dP} = 4.5z_k^2/b_k \quad (158)$$

and the electrical contribution to  $\Phi_{v2}^0$  is

$$(\Phi_{v2}^0)_e = \Sigma_i 4.5\nu_i z_i^2/b_i \quad (159)$$

If  $d \ln D/dP$  may be determined from Redlich's value of the limiting slope of  $\Phi_{v2}$  vs.  $\sqrt{I/2}$ , the numerical factor is changed from 4.5 to 4.8.

#### Electrolyte Solubility in Different Solvents

The interaction of an electrolyte and a non-electrolyte is most simply considered from equation 130 in the limit of zero salt concentration.

$$\mu_{ks} - \mu_{ks}^0 = (N\epsilon^2 z_k^2/2b_k)(1/D - 1/D_0) \quad (160)$$

The electrostatic contribution to the chemical potential of an ion should be a linear function of the reciprocal dielectric constant. This is often called the Born equation<sup>28</sup>. If there is no deviation from the laws of ideal solution except as indicated in equation 160, the solubility, corrected to zero ionic strength by equation 138, is

$$\ln \frac{N_s}{N_{s0}} = -\Sigma_i \nu_i (\mu_{is} - \mu_{is}^0)/RT \Sigma_i \nu_i = (N\epsilon^2 z_+ z_-/2b_s)(1/D - 1/D_0) \quad (161)$$

Bjerrum<sup>29</sup> was the first to note that non-electrostatic effects must also be taken into account and that equation 161 should give the logarithm of the solubility ratio of the salt minus that for a non-electrolyte molecule having the average properties of the ions, such as argon for potassium or chloride ion and neon for sodium or fluoride ion. The non-electrostatic effects are particularly large for large organic ions.

Debye<sup>20</sup> has noted that equation 161 should give only a poor approximation in the case of mixed solvent because it ignores the sorting of the solvents by the ions which attract the molecules of high dielectric constant more than those of low. Debye's calculations treat the solvent as a homogeneous structureless medium, and are still far from satisfactory as a representation of the experimental facts.<sup>30</sup>

#### Salting-Out of Non-Electrolytes

For solutions which are dilute in the non-electrolyte as well as in electrolyte, we may write

with  $\delta_n$  a constant.

$$1/D = 1/D_0[1 + (\delta_n/D_0)C_n] \cong 1/D_0 - (\delta_n/D_0^2\rho_0)m_n \quad (163)$$

For a single non-electrolyte and single salt, this leads to

$$\ln \gamma_{\pm} = (N\epsilon^2 z_+ z_- \delta_n / 2RTb_s D_0^2 \rho_0) m_n \quad (164)$$

$$\ln \gamma_n = (N\epsilon^2 z_+ z_- \delta_n / 2RTb_s D_0^2 \rho_0) \nu m_s \quad (165)$$

$$\varphi = x_s \varphi_s + x_n \varphi_n + (N\epsilon^2 z_+ z_- \delta_n / 2RTb_s D_0^2 \rho_0) M x_s x_n \quad (166)$$

in which  $M = \nu m_s + m_n$ ,  $x_s = \nu m_s / M$ ,  $x_n = m_n / M$ ,  $\varphi_s$  is the value of  $\varphi$  in a solution of salt alone at the concentration  $m_s$ , and  $\varphi_n$  is the value of  $\varphi$  in a solution of the non-electrolyte alone at the concentration  $m_n$ . Equations 160 and 164 may be used to calculate the effect of a non-electrolyte on salt solubility. If the dielectric constant is diminished by the non-electrolyte,  $\delta_n$  is negative; since  $z_-$  is also negative,  $\ln \gamma$  is increased and the solubility is diminished. Equation 165 may be used to calculate the effect of the electrolyte on the non-electrolyte. If  $\delta$  is negative, this solubility is decreased by the addition of salt, or the non-electrolyte is salted out. We may call the term in parenthesis in these equations the "salting-out constant." Equation 166 may be used to determine this salting-out constant from freezing point depressions or other measurements of the potential of the solvent. The salting-out constant depends upon the temperature, upon the dielectric constant of the solvent, upon the rate of change of dielectric constant with changing concentration of the non-electrolyte solute, and upon the valences and radii of the ions. So it depends upon specific properties of both the solutes.

The approximations made in the derivation of these equations indicate that deviations from the simple linear equations might be expected at low concentrations. For non-electrolytes which lower the dielectric constant, however, the linearity persists to high concentration, particularly for equation 165. The decrease of the dielectric constant of water by most solutes may be largely accounted for by the displacement of water by a substance of much lower dielectric constant. This displacement also occurs in solutions of ions and dipolar ions, and these solutions do appear to show a salting-out in addition to the salting-in due to their electrostatic effects. The two effects can be distinguished to a certain extent because the salting-out remains proportional to the concentration while the electrostatic effect increases much less rapidly at higher concentrations. Probably the chief difficulty is in those effects which would remain if the ionic charges were removed, which we have indicated by  $\delta_2$  in the charging process. Most



accounted for largely by the difficulty of separating the very cohesive water molecules, and it therefore depends largely upon the volume of the solute molecules. The simplest theory leads to a negative term in  $B_{ij}$  of equation 31 or 146 which is proportional to the product of the molecular volumes of the  $i$ 'th and the  $j$ 'th species. Even for the larger monatomic anions this effect appears to be larger than the salting-out effect<sup>31</sup>. The Hofmeister series of the ions is largely dependent upon the differences in the salting-out effect and this non-electrostatic effect; for spherical ions the first is proportional to  $1/b$  and the second to  $b^3$ . In some of the Hofmeister series more specific effects are doubtless involved.

#### "Specific Ion Interaction"

In an electrolyte solution the forces of repulsion between like charges will usually keep ions of the same charge separated sufficiently that the shorter range interactions between ions of the same sign may be ignored. This is a statement of Brønsted's principle of "specific ion interaction,"<sup>32</sup> It leads to certain conclusions concerning the behavior of solutions of mixed salts which have been verified experimentally, and which lead to important simplifications in the treatment of such mixtures. It also gives the basis for using a single value of  $a$  in solutions containing only two ion species<sup>33, 31</sup> The important distance is that between the cation and the anion, and the error is not large in attributing this distance to a pair of cations or to a pair of anions. If the cations are smaller there will be a shell around each cation from  $2b_+$  to  $a = (b_+ + b_-)$  into which another cation can enter though we assume that it cannot; however, the electrostatic repulsion will prevent many cations from entering this shell. Around each anion there is a shell from  $(b_+ + b_-) = a$  to  $2b_-$  into which no other anion can enter, though we assume that it can. However, we have assumed electrostatic repulsion, so we have not assumed that many ions enter this shell. The two errors also tend to compensate. The principle of specific ion interaction also indicates that the salting-out effect and the non-electrostatic effect should also be restricted almost entirely to reactions between oppositely charged ions or between ions and neutral molecules.

#### Electrostatic Association

The importance of that part of equation 106 which we have neglected also depends upon the nearness of approach of oppositely charged ions. The "higher terms" of the series expansion of equation 107 have been treated by Bjerrum, by Müller, by Gronwall and La Mer, and by Fuoss

and Kraus<sup>34, 35, 36, 37</sup>. Bjerrum's treatment, which has been extended by Fuoss and Kraus, is the most enlightening. If equation 103 be applied to obtain the number of ions per unit thickness of a spherical shell at the limit of zero ionic strength, it yields

$$n'_k = 4\pi r^2 c_k = 4\pi r^2 c_k^0 e^{-\psi_{ik}/kT} = 4\pi r^2 c_k^0 e^{-\epsilon^2 z_i z_k / kTD r} \quad (167)$$

This function,  $n'_k$ , passes through a minimum at

$$r = -\epsilon^2 z_i z_k / 2kTD = a'. \quad (168)$$

For two ions with charges of the same sign the minimum occurs at a negative distance and has no physical significance. The number of ions near a given ion of the same charge is always less than the number in an equal volume at a greater distance. If the charges have opposite signs, however, the distance  $a'$  is positive, and the number within a spherical shell increases rapidly as the distance decreases beyond  $a'$  until it suddenly drops to zero at the distance  $a$  because of the mechanical repulsion.

The number of ions at or near the minimum distance is so small that it is permissible to neglect them and to divide the ions sharply into two classes, those which are at distances smaller than  $a'$  and those which are at larger distances. In sufficiently dilute solutions the chance that more than one ion is at a smaller distance is negligibly small, and all other ions will be at a distance so very much larger that they may be considered as equidistant from this ion and the central ion. This is equivalent to saying that the two have associated to form a complex of valence  $(z_i + z_k)$ . Bjerrum calculates the constant for this association from the total number at a distance larger than  $a$  but smaller than  $a'$  as

$$K = \frac{1}{C} \int_a^{a'} n'_k dr = \frac{4\pi N}{1000} \int_a^{a'} e^{-\epsilon^2 z_i z_k / kTD r} r^2 dr \quad (169)$$

since the concentration in moles per liter is  $C = 1000c_k^0/N$ . For purposes of integration the variable is changed to  $y = -\epsilon^2 z_i z_k / kTD r = 2a'/r$  to give

$$K = 0.032\pi N a'^3 \int_2^{2a'/a} e^y y^{-4} dy \quad (170)$$

Bjerrum extends to higher concentrations by calculating from the Debye-Hückel equation (the first of equation 138) with size  $a'$  rather than  $a$  activity coefficients for the unassociated ions and for the complex if its charge is not zero. The effect is neglected if  $a'$  is less than  $a$ , that is for large ion pairs of the opposite sign and for all pairs of the same sign. When

$z_i \neq z_k$  or when there are more than two ion species, a constant is calculated for each type of association. The fact that these constants may be calculated from electrostatic theory and ion sizes and valences shows that the association is very different from that of weak electrolytes such as the carboxylic acids. The effect on the thermodynamic properties is the same, however, so that it is impossible to be sure that a measured association is electrostatic because it leads to a reasonable size by Bjerrum's method of calculation. If the calculated size is not reasonable, on the other hand, we may be sure that part of the effect, at least, is not electrostatic.

The methods of Müller<sup>35</sup> and of Gronwall and La Mer<sup>36</sup> of determining the "higher term" correction are more mathematical and do not lead to an arbitrary size  $a'$ , above which the correction is neglected. However, calculated by the other methods, the correction is small for sizes larger than the  $a'$  of Bjerrum's method.

#### Güntelberg-Müller Charging Process

If the effect of the ions on the volume and on the dielectric constant is neglected, it is simpler to obtain the first term of equation 138 by another charging process found independently by Güntelberg<sup>38</sup> and by Müller<sup>35</sup>. Instead of transferring all the ions, the transfer is limited to  $\nu_+$  positive ions and  $\nu_-$  negative ions. So the second charging process is carried out at constant  $\kappa$ . Then equation 126 is replaced by

$$\begin{aligned} W'_e &= \sum_i \nu_i \int_{\lambda=0}^1 (\epsilon^2 z_i^2 / D_0) [1/b_i - \kappa_0 / (1 + \kappa_0 a)] \lambda d\lambda \\ &= \sum_i (\nu_i \epsilon^2 z_i^2 / 2D_0) [1/b_i - \kappa_0 / (1 + \kappa_0 a)] \end{aligned} \quad (171)$$

and

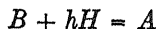
$$W'_e - W_e^0 = -\sum_i (\nu_i \epsilon^2 z_i^2 / 2D_0) \kappa_0 / (1 + \kappa_0 a) \quad (172)$$

But  $W'_e - W_e^0$  is  $\mu_{se} - \mu_{se}^0$ , and  $\sum_i \nu_i z_i^2 = -\nu_+ z_+ z_-$ , so equation 172 is the same as the first term of equation 138. This charging process gives a somewhat different answer from that obtained with the Debye charging process for the "higher terms" in the expansion of the Boltzmann exponential. Neither need be right for the conditions of integrability are not fulfilled for these higher terms. The Güntelberg-Müller charging process is particularly useful for the study of the effect of symmetrical ions on unsymmetrical ions or on dipolar ions, for which no other treatment has been found, and for which the treatment has been limited to the first two terms of the series expansion where there is no difficulty because of the conditions of

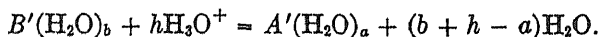
In water solutions at room temperature,  $a'$  is  $-3.5z_+z_-$  Ångstroms. The electrostatic association is not large for any univalent ions, and it is quite negligible for the amino acid ions. We might expect a measurable effect for protein ions with large charges, but not with those which exist in appreciable concentrations near the isoelectric point. In all these cases we might expect the effect of "higher terms" to be much smaller than the effects of the lack of spherical symmetry which are discussed in chapter 12.

### Chemical Equilibria

We may now return to the equilibria in which chemical reactions are involved. We shall be particularly interested in homogeneous reactions of the type



This includes the formation of any binary compound, for which  $h = 1$ , and also the acid-base equilibrium for any polyvalent acid, base or ampholyte, for which  $H$  is the hydrogen ion and  $h$  may have any positive, integral value. For this case its application should be limited to solutions containing sufficient water to make unimportant the formation of water by the reaction



Equation 37 becomes

$$\ln (A)/(B)(H)^h = \ln K' = \ln K + \ln \gamma_B \gamma_H^h / \gamma_A \quad (173)$$

in which  $(A)$ ,  $(B)$  and  $(H)$  are the concentrations of the corresponding species. If the concentrations  $(A)$ ,  $(B)$  and  $(H)$  are not expressed as mole fractions, there will be factors in each  $\gamma$  to correct to these units. These will cancel in  $\gamma_A$  and  $\gamma_B$ . If the concentrations are expressed in moles per liter of solution, the correction for an ideal solution will be

$$\ln K'_I/K = h \ln (\gamma_H)_I = h \ln (V_N/V_N^0) \quad (174)$$

in which  $V_N$  is  $V$ , the total volume, divided by  $\sum n_i$ , the total number of moles including those of the solvent, and  $V_N^0$  is  $V_N$  for the standard state. If the concentrations are expressed in moles per kilogram of solvent,  $V_N$  is replaced by the mass of solvent in one total mole of solution. For an infinitely dilute solution,  $V_N$  may be replaced by the average molecular weight of the solvent. This correction has almost always been made for calculations of chemical equilibrium, but almost never for calculations of rates of reactions although these calculations are also based on equation

In some ways the treatment of the last term of equation 173 is simpler than the treatment of its component parts,  $\ln \gamma_A$ , etc. The volume of  $B + hH$  is not very different from that of  $A$ , and effects which depend upon displacement of the solvent do not depend much upon the extent of the reaction. Conversely, the extent of the reaction, and therefore  $\ln K'$ , does not vary much because of non-electrolyte reactions or salting-out. Large changes in equilibria appear to result from changes in ionic charges, or in dipole moments if the moments are very large as in dipolar ions, or the dipole is very near the surface as in the hydroxyl, amino or other groups which lead to hydrogen bonds, or sometimes in moments higher than the dipole if these are very large. We shall confine our attention to the effects which have been discussed in this chapter.

If we assume that the ions are all spherical and with the same collision diameter  $a$ ,

$$\ln K' - \ln K = \ln \gamma_B \gamma_H / \gamma_A =$$

$$\frac{\epsilon^2}{2RTD} \left[ \left( 1 - \frac{D}{D_0} \right) \left( \frac{z_B^2}{b_B} + \frac{hz_H^2}{b_H} - \frac{z_A^2}{b_A} \right) - \frac{\kappa}{1 + \kappa a} (z_B^2 + hz_H^2 - z_A^2) \right] \quad (175)$$

Since

$$z_A = z_B + hz_H \quad (176)$$

$$z_B^2 + hz_H^2 - z_A^2 = -[2hz_Bz_H + h(h-1)z_H^2] \quad (177)$$

If  $h = 1$ , this becomes

$$z_B^2 + hz_H^2 - z_A^2 = -2z_Bz_H \quad (178)$$

If  $B$  and  $H$  are spheres and  $h = 1$ , it is probably more accurate to look upon  $A$  as made up of two spheres separated by a distance  $r$ . The activity coefficient for such a model may be obtained by assuming that

$$\ln \gamma_B \gamma_H / \gamma_A = \ln C_r / C_\infty \quad (179)$$

in which  $C_r$  is the concentration of  $H$  at a distance  $r$  from a  $B$  molecule and  $C_\infty$  is the concentration at an infinite distance. If  $r$  is not less than  $a$ , equations 103 and 117 yield

$$\ln \gamma_B \gamma_H / \gamma_A = \frac{-\epsilon^2 z_B z_H}{RT r} \left[ \frac{e^{-\kappa(r-a)}}{D(1 + \kappa a)} - \frac{1}{D_0} \right] \quad (180)$$

The expression does not vary much with  $a$  when  $a$  is approximately equal to  $r$ , and it is a sufficiently good approximation to take  $a = r$ , which yields

This effect of changing ionic strength is only a crude approximation for the two sphere model because the mutual interaction of the two ion atmospheres is neglected<sup>32, 40</sup>. The error does not affect the limiting law, but it does become more serious the higher the salt concentration. This effect is the same as that in equation 175 for spheres with equal collision diameters. The effect of changing dielectric constant is not the same for equations 175 and 181. Either gives correctly the effect for the corresponding model in a continuous medium of uniform dielectric constant, and the two may be made to give the same numerical result by the appropriate choice of parameters. It is probably more accurate to use for  $A$  the ellipsoidal model described in Chapter 12.

We are often interested in reactions involving the hydrogen ion when the hydrogen ion activity is assumed to be determined. If we let  $H$  represent the hydrogen ion, we may obtain from equation 173

$$\ln (A)/(B) = \ln K'' + h \ln a_H = \ln K + h \ln a_H + \ln \gamma_B/\gamma_A \quad (182)$$

$K''$  differs from  $K'$  in that it lacks the activity coefficient of the hydrogen ion. The proton is so small that in this case it is probably permitted to assume that  $A$  and  $B$  have the same size and shape if they are hydrated to the same extent, and we will assume them to be spheres of radius  $b$  and collision diameter  $a$ . Then by equation 175

$$\ln K'' - \ln K = \ln \gamma_B/\gamma_A = -\frac{\epsilon^2}{2RTD} h(2z_B + h) \left[ \frac{1}{b} \left( 1 - \frac{D}{D_0} \right) - \frac{\kappa}{1 + \kappa a} \right] \quad (183)$$

The usefulness of this equation for calculating the effect of changing medium is limited by the difficulty of determining  $a_H$  in any but dilute aqueous solutions. This difficulty arises both from our inability to interpret measurements of the electromotive force of a cell with a liquid junction between two solvents and from the fact that the hydrogen ion does not exist in solution as a simple proton,  $H^+$ , but is solvated. If there is any water present, the hydrogen ion probably exists largely as  $H_3O^+$ . The ratio of two constants is conveniently measured, however, if  $h$  is unity for each. Then

$$\ln \frac{(A_1)(B_2)}{(A_2)(B_1)} = \ln \frac{K_1''}{K_2''} = \ln \frac{K_1}{K_2} - \frac{\epsilon^2}{2RTD} \left[ \left( \frac{2z_{B_1} + 1}{b_1} - \frac{2z_{B_2} + 1}{b_2} \right) \left( 1 - \frac{D}{D_0} \right) - 2(z_{B_1} - z_{B_2}) \frac{\kappa}{1 + \kappa a} \right] \quad (184)$$

The effect of changing ionic strength is given by equation 183 or by equa-

A better fit may be obtained empirically by adding a term proportional to the salt concentration similar to that in equation 140, but calculation of the constant from theory depends upon the difference in the effects of  $A$  and  $B$  on the dielectric constant as well as on the difference in their valences, and it would probably involve a serious error if this were calculated on the assumption that the charges are spherically symmetrical in each. The existence of this term may limit the application of equation 183 or 184 to rather dilute salt solutions.

For a reaction involving only one hydrogen ion, and within the limits of the applicability of equation 183 with  $D = D_0$ , the effect of increasing ionic strength will be to increase the ratio  $(A)/(B)$  if  $z_B$  is zero or positive, but to decrease this ratio if  $z_B$  is negative. This is important in the effect of salts on indicators of various valence types. Conversely, if a mixture of  $A$  and  $B$  is used as a buffer, increasing ionic strength at a constant ratio  $(A)/(B)$  will decrease the hydrogen ion concentration if  $z_B$  is zero or positive, but will increase the hydrogen ion concentration if  $z_B$  is negative. In both these cases, the magnitude of the effect increases rapidly with the absolute value of  $z_B$ .

It is convenient for some purposes to make the comparison always with the neutral molecule by allowing  $h$  to take negative integral values as well as positive and taking  $z_B = 0$ . The equation 183 becomes

$$\ln K'' - \ln K = -\frac{\epsilon^2 h^2}{2RTD} \left[ \frac{1}{b} \left( 1 - \frac{D}{D_0} \right) - \frac{\kappa}{1 + \kappa a} \right] \quad (185)$$

In dilute solutions we may assume that  $D = D_0$  and that

$$\ln K'' - \ln K = \frac{\epsilon^2 h^2}{2RTD} \frac{\kappa}{1 + \kappa a} \quad (186)$$

The right hand side is intrinsically positive, so  $\ln K''$  increases with increasing salt concentration. In a titration curve the amount of acid or base which reacts with a given amount of isoelectric ampholyte at a given hydrogen ion concentration should increase with increasing ionic strength, and the magnitude of the change should be greater as the distance from the isoelectric point increases. In solubility measurements we are interested in the activity coefficient of the component  $B$ , which will change with salt concentration because the fraction of the substance in unionized form varies even if the species  $B$  behaves as an ideal solute. The activity coefficient of  $B$  will vary as though it were an ion of valence  $\sqrt{[\Sigma h^2(A_h)]/[\Sigma(A_h)]}$ <sup>41</sup>. Since  $\Sigma(A_h)$  is the total concentration of the

is obvious that the contribution to the ional concentration corresponds to this valence. Making the same assumptions as in equation 185, the excess, or non-ideal chemical potential per mole of the  $h$ 'th species is

$$\mu_h^e = \frac{\epsilon^2}{2D} \left[ \frac{1}{b} \left( 1 - \frac{D}{D_0} \right) - \frac{\kappa}{1 + \kappa a} \right] h^2 \quad (187)$$

The excess chemical potential per mole of the component is

$$\mu_B^e = [\Sigma_h(A_h) \mu_h^e] / [\Sigma_h(A_h)] \quad (188)$$

and the activity coefficient is given by

$$\ln \gamma_B = \mu_B^e / RT = \frac{\epsilon^2}{2RTD} \left[ \frac{1}{b} \left( 1 - \frac{D}{D_0} \right) - \frac{\kappa}{1 + \kappa a} \right] [\Sigma_h h^2(A_h)] / [\Sigma_h(A_h)] \quad (189)$$

which is equal to  $\ln \gamma$  for an ion whose valence is the square root of the last factor. The effective valence increases with increasing salt concentration, and this will make  $a$  appear smaller than for a single ion  $A_h$ .

For simple amino acids the effect is not large. If  $h$  can be only 0, +1, or -1, the apparent valence at the isoelectric point is  $2\sqrt{K_+K_-}$ , which is approximately  $2 \times 10^{-4}$  for the simple amino acids.  $K_- = K_2$  as the constants are usually written and  $K_+ = 1/K_1$ . For the proteins, however, there are so many more dissociating groups, and their constants are so much nearer each other that the apparent valence may be great enough to show experimentally a variation of  $\ln \gamma_B$  which is linear in the square root of the ionic strength. This effective valence may be determined from a titration curve. The total concentration of protein, which must be independent of the hydrogen ion concentration, is

$$\Sigma_h(A_h) = \Sigma_h K'_h(B)(H^+)^h \quad (190)$$

$$\begin{aligned} \frac{d \Sigma_h(A_h)}{d \ln (H^+)} &= \Sigma_h h K'_h(B)(H^+)^h + K'_h(B)(H^+)^h \frac{d \ln (B)}{d \ln (H^+)} \\ &= \Sigma_h A_h \left[ h + \frac{d \ln (B)}{d \ln (H^+)} \right] = 0 \end{aligned} \quad (191)$$

$$\frac{d \ln (B)}{d \ln (H^+)} = - \frac{\Sigma_h h(A_h)}{\Sigma_h(A_h)} \quad (192)$$

The change in the net charge is

$$\frac{d \Sigma_h h(A_h)}{d \ln (H^+)} = \Sigma_h \left[ h^2 K'_h(B)(H^+)^h + h K'_h(B)(H^+)^h \frac{d \ln (B)}{d \ln (H^+)} \right] \quad (193)$$



and the change in net charge per mole of protein is

$$\frac{1}{\Sigma_h(A_h)} \frac{d \Sigma_h h(A_h)}{d \ln(H^+)} = \frac{\Sigma_h h^2(A_h)}{\Sigma_h(A_h)} - \left[ \frac{\Sigma_h h(A_h)}{\Sigma_h(A_h)} \right]^2 \quad (194)$$

Since the solution must be electrically neutral,

$$\frac{\Sigma_h h(A_h)}{\Sigma_h(A_h)} = a - b + \frac{(\text{OH}^-) - (\text{H}^+)}{\Sigma_h(A_h)} \quad (195)$$

if  $a$  moles of acid and  $b$  moles of base are added per mole of protein. The necessary conditions are that the volumes of the solution and the activity coefficients remain constant during the titration and the titrating acid and base are completely ionized. They are met approximately in an ordinary titration. The effective valence of the albumins at the isoelectric point appears to be between two and three.

The effect on  $\ln \gamma$  should affect the solubility measurements which are discussed in Chapters 23 and 24. If the saturating body is the neutral ampholyte

$$\ln S/S_0 = -\ln \gamma_B \quad (196)$$

If the saturating body is a salt  $AX_x$ , in which each ampholyte ion is combined with  $x$  ions of valence  $z_x$ , so that the valence of the ampholyte in the saturating body is  $-xz_x$ , the solubility is given by

$$\ln S/S_0 = -\frac{1}{1+x} (\ln \gamma_B + x \ln \gamma_x) \quad (197)$$

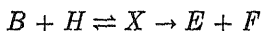
In equations 196 and 197  $S$  is the solubility, preferably as mole fraction, and  $S_0$  is the solubility at the same pH but zero ionic strength.

### Chemical Kinetics

The treatment of reaction rates is very similar to that of chemical equilibrium. We assume that two or more molecules must approach to a certain distance and with a certain orientation before they can react. The fact that many reactions proceed at a measurable rate which increases rapidly with the temperature indicates that only the very small fraction of such pairs react which have an energy very much larger than the average. Probably this energy must be associated with one definite mode of vibration. Except for very simple molecules we are not able to calculate absolute reaction rates, but progress has been made in the study of the effect of media upon these rates. The group of molecules in the proper position and with sufficient energy to react is called the "critical complex"

called the "energy of activation". It is usually assumed that the energy necessary for reaction is  $\ddagger$  constant for any reaction. The energy of activation then varies with the temperature only through the relatively slight variation of the average energy and it varies not at all with the composition.

Given the reaction



in which  $X$  is the critical complex, it is customary to assume that almost every molecule of  $X$  formed from  $B$  and  $H$  goes instantly to  $E$  and  $F$ , and that the number of molecules of  $E$ , or  $F$ , formed in unit time is proportional to the very small number of molecules of the critical complex. It is further assumed that this critical complex can be treated as though it were in thermodynamic equilibrium with the reactants  $B$  and  $H$ . One justification of this assumption is that there are a large number of complexes, which we would designate as  $A$ , which differ very little from  $X$ , but which do not react.

The first application of the theory was made by Brønsted<sup>42</sup> to the effect of electrolytes upon the rates of reactions between ions. He explained why salts increase the rate of some reactions, decrease those of others, and have almost no effect upon a third group. Brønsted's method is to calculate the rate as

$$-d(B)/dt = k_1(X) = k(B)(H)\gamma_B\gamma_H/\gamma_X \quad (198)$$

We may define the ordinary rate,  $k'$ , as

$$-d(B)/dt = k'(B)(H) \quad (199)$$

and

$$k' = k\gamma_B\gamma_H/\gamma_X \quad (200)$$

If  $\gamma_X$  is the same as  $\gamma_A$  for a stable complex of the same configuration  $\ln(k'/k)$  may be determined from equation 181. Usually  $h = 1$ , so the second term of 177 becomes  $-2z_Bz_H$ , but  $\ln(k'/k)$  may probably be determined more accurately from equation 181.<sup>39, 43</sup> From either of these equations it is obvious that an increase of the ionic strength increases the rate of reaction if  $B$  and  $H$  have charges of the same sign, decreases the rate if their charges have opposite signs, and has no effect, as far as these equations go, if either charge is zero. This accounts qualitatively for Brønsted's results. The quantitative expression differs a little from the original one of Brønsted which antedated the theory of Debye and

Hückel. It is also clear that an increase of dielectric constant has an effect of the same sign as an increase of the ionic strength. This effect is linear in the reciprocal of the dielectric constant and depends upon the distance of approach  $r$ . Reasoning such as that which leads to equation 181 can give a rough qualitative picture of the variation of the rate of reaction between an ion and a neutral molecule with varying ionic strength and with varying dielectric constant<sup>43</sup>, but no very definite results can be reached beyond those of equation 181 or 175, or the more accurate models of Chapter 12. There is one case which is very interesting although it is difficult to establish that any particular reaction is of this type. It is the case of neutral reactants and an ionic or dipolar critical complex, and  $\ln(k'/k)$  differs from equation 181 by  $\ln \gamma_B \gamma_H$ <sup>39, 44</sup>.

Effects which we usually describe as catalytic are probably of quite a different nature from those described in equation 181, although either changing dielectric constant or changing ionic strength may change the rate at least several hundred fold. A catalyst usually forms a new critical complex with the reactants, which often, although not necessarily, has a much smaller energy of activation than the critical complex without the catalyst. We are not yet in position to explain either the great catalytic effect or the marked specificity in enzyme reactions.

The material covered in this chapter is so extensive that it has been necessary to give the thermodynamics very briefly, except for a few sections of particular interest in protein chemistry, and to limit the physical theory to the simplest possible model in moderately dilute solutions. The treatment of this model serves as an illustration of our methods of developing physical theories, and it does give a surprisingly good approximation of the behavior of simple ions and non-electrolytes. The treatment of amino acids in the chapters which follow requires one important extension of the simple theory. The distribution of the charges within dipolar ions must be taken into account in the study of their interactions with ions, with simple non-electrolytes, and with each other. This and more complicated models will be treated theoretically in Chapter 12.

<sup>44</sup> Stearn, A. E., and Eyring, H., *J. Chem. Physics*, 5, 113 (1937).

## Chapter 4

### Dipolar Ions and Acid-Base Equilibria

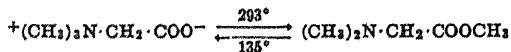
BY JOHN T. EDSALL

The dipolar ions considered in this book may be formed, by intramolecular proton shifts, from uncharged isomers of lower dipole moment. Thus glycine may exist either as the dipolar ion,  $^+\text{H}_2\text{N}\cdot\text{CH}_2\cdot\text{COO}^-$ , or as the "uncharged molecule,"  $\text{H}_2\text{N}\cdot\text{CH}_2\cdot\text{COOH}$ , and either of these forms may go over into the other through interchange of a proton between the carboxyl and the amino group<sup>1</sup>.

The position of the equilibrium between the dipolar ion and the uncharged molecule depends on the acidic and basic properties of the groups involved. Thus the aliphatic amino acids in water exist almost entirely in the dipolar ionic form; the amino-benzoic acids as a mixture of the two forms, both in considerable amount; and the aminophenols almost entirely in the uncharged form. To obtain a basis for predicting the behavior of any particular type of compound, the nature of the acid-base equilibria involved must be more precisely formulated.

Probably the most satisfactory basis for such a formulation is Brönsted's<sup>2</sup> theory of acids and bases. According to Brönsted's conception, an acid is any substance capable of giving up a proton, such as  $\text{CH}_3\text{COOH}$ , the  $\text{NH}_4^+$  ion, or the ethylene diammonium ion,  $^+\text{H}_2\text{N}\cdot\text{CH}_2\cdot\text{CH}_2\cdot\text{NH}_3^+$ . A base is any substance capable of taking up a proton, such as the  $\text{CH}_3\text{COO}^-$  ion, ammonia, or the hydroxyl ion. For every acid (A) there is a con-

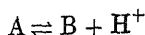
<sup>1</sup> In the case of the betaines and their isomers, it is a charged alkyl group which is interchanged rather than a proton. For example, betaine (N-trimethyl glycine) and its isomer, dimethyl glycine methyl ester, are both stable compounds at room temperature. The ester when heated above its boiling point at 135°, undergoes an intramolecular rearrangement and passes over into the betaine. The latter melts at 293°, and on melting undergoes the reverse transformation into the ester.



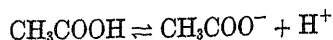
Similar relations are known for a number of other betaines (See R. Willstätter, *Ber.* 35, 584 (1902); R. Willstätter and W. Kahn, *Ber.* 37, 401 (1904)).

The mechanism of these intramolecular rearrangements, and the magnitude of the free energy changes involved at various temperatures, are subjects calling for investigation. The betaines and the isomeric esters

jugate base (B), obtained from it by the removal of a proton, according to the reaction



Thus we may write for acetic acid



its acid dissociation constant being

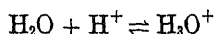
$$K_A = \frac{(CH_3COO^-)(H^+)}{(CH_3COOH)} \quad (1)$$

Similarly for the ammonium ion

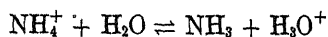
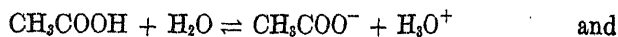


$$K_A = \frac{(H^+)(NH_3)}{(NH_4^+)} \quad (2)$$

These equations are incomplete in one respect. The proton cannot exist as such in appreciable quantity in any solution. Immediately upon being liberated from an acid, it must combine with one of the bases present. In aqueous solutions the solvent water acts as a base:



The reactions given above may, therefore, be written more precisely



In dilute aqueous solution, however, we may use expressions such as equations 1 and 2 above, remembering that  $(H^+)$  really denotes the activity of the  $H_3O^+$  (hydronium) ion; while the water molecule, which plays an essential part in the acid-base equilibrium, is omitted from the mass law equation because its concentration is so large as to be essentially constant throughout the reaction. We shall, therefore, employ dissociation constants written in the same form as 1 and 2 throughout the present discussion.

According to this definition, an acid or base may carry an electric charge which is positive, zero, or negative. Table 1, giving some acids and their conjugate bases, illustrates this point.

It is clear that a given molecule or ion may function both as an acid and as a base; and that a given acid carries a charge always one unit more positive (or less negative) than that of the conjugate base.

In general, individual acidic groups within a molecule are of two

1 gives rise to negatively charged, type 2 to uncharged basic groups. The total charge of any molecule or ion depends on the number and state of the acidic or basic groups which it carries.

Brönsted's treatment is immediately applicable to acid-base equilibria in any solvent. During the first part of this chapter, however, we shall confine our attention to dilute aqueous solutions, subsequently considering the important influence of variation in the solvent on the equilibrium between dipolar ions and uncharged molecules.

To be capable of forming a dipolar ion, a substance must contain at least two acidic groups in strongly acid solution; at least one group must be of the uncharged type and at least one of the cationic type; that is, the molecule must be an amphoteric substance or ampholyte. Furthermore, if the substance is to exist in the isoelectric state as a dipolar ion, not as an uncharged molecule, the acidity of the uncharged acid group must be greater than that of the cationic group.

These general considerations may be illustrated by the particular case of glycine, the simplest of the aliphatic amino acids. In strongly acid

TABLE 1

Acid	Net Charge	Base	Net Charge
$^+\text{H}_3\text{N}\cdot\text{CH}_2\cdot\text{CH}_2\cdot\text{NH}_3^+$	+2	$^+\text{H}_3\text{N}\cdot\text{CH}_2\cdot\text{CH}_2\cdot\text{NH}_2$	+1
$^+\text{H}_3\text{N}\cdot\text{CH}_2\cdot\text{CH}_2\cdot\text{NH}_2$	+1	$\text{H}_2\text{N}\cdot\text{CH}_2\cdot\text{CH}_2\cdot\text{NH}_2$	0
$\text{HOOC}\cdot\text{COOH}$	0	$\text{HOOC}\cdot\text{COO}^-$	-1
$\text{HOOC}\cdot\text{COO}^-$	-1	$^-\text{OOC}\cdot\text{COO}^-$	-2

solutions, glycine exists in the form of the cation  $^+\text{H}_3\text{N}\cdot\text{CH}_2\text{COOH}$ . On the addition of a base, such as hydroxyl ion, both the carboxyl and the ammonium group tend to give up a proton to the base. Which will do so more readily?

The acidic dissociation constant of the carboxyl group in acetic acid is approximately  $10^{-4.8}$ , and the presence of the adjoining positively charged  $-\text{NH}_3^+$  group in the glycine cation might be expected to repel a proton from the carboxyl group and thereby increase acid dissociation. The acid dissociation constant of the methylammonium ion on the other hand is about  $10^{-10.7}$ , nearly a million times weaker than that of the carboxyl group. The first dissociation constant of glycine,  $K_1$ , is found to be

$$K_1 = \frac{(\text{H}^+)(\text{R})}{(\text{R}^+)} = 10^{-2.3}$$

where  $(\text{R}^+)$  stands for the concentration of glycine cation, and  $(\text{R})$  for that of isoelectric glycine. This value of  $10^{-2.3}$  is comprehensible as repre-

it should be due to the ionization of the ammonium group, for this dissociation constant of glycine is nearly three hundred million times as great as that of the methyl-ammonium ion. The presence of the adjoining carboxyl group could not account for such a gigantic effect. Hence isoelectric glycine must contain an ionized carboxyl group, and is, therefore, a dipolar ion.

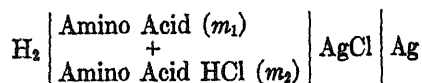
This argument is essentially identical with that originally employed by Adams<sup>3</sup> and Bjerrum<sup>4</sup> to prove the dipolar ionic structure of the aliphatic amino acids. It remains one of the most powerful and convincing of all the arguments in favor of this structure.

#### Thermodynamic Dissociation Constants of the Amino Acids

To make these considerations more precise, we may now consider some very exact values for the dissociation constants of certain amino acids at various temperatures. These have been determined by measurement of the electromotive force of concentration cells without liquid junction. For the determination of the first dissociation constant

$$K_1 = \frac{(H^+)(^+H_2N \cdot R \cdot COO^-)}{(^+H_2N \cdot R \cdot COOH)} \quad (3a)$$

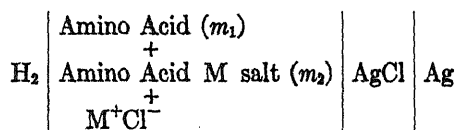
(quantities in brackets refer to activities)  
cells of the type:



were employed. For the measurement of the second dissociation constant,

$$K_2 = \frac{(H^+)(H_2N \cdot R \cdot COO^-)}{(^+H_2N \cdot R \cdot COO^-)} \quad (3b)$$

the cells were of the type:



(where  $M^+$  is generally  $Na^+$  or  $K^+$ )

The principles of the method are due to Harned<sup>5</sup> and will not be treated here. This method is capable of far greater accuracy than ordinary measurements on the hydrogen, quinhydrone or glass electrodes which involve liquid junctions. Thus Harned and Ehlers<sup>6</sup> have determined the

dissociation constant of acetic acid in water at 25° as  $1.754 \times 10^{-5}$  (pK = 4.756), while MacInnes and Shedlovsky<sup>7</sup> have obtained from very accurate conductivity measurements the value  $1.753 \times 10^{-5}$ . The pK

TABLE 2. CONSTANTS IN THERMODYNAMIC EQUATIONS FOR AMINO ACIDS (0° TO 60°C.)

Calculated from the data of B. B. Owen and P. K. Smith, A. C. Taylor and E. R. B. Smith by R. A. Robinson and E. R. B. Smith, from the equations of H. S. Harned and R. A. Robinson, *Trans. Faraday Soc.*, **36**, 973 (1940).

$$\begin{aligned} \text{pK} &= \frac{A'}{T} - C' + D'T \\ \Delta F^\circ &= A - CT + DT^2 = -RT \ln K \\ \Delta H^\circ &= A - DT^2 \\ \Delta C_p^\circ &= -2DT \\ \Delta S^\circ &= C - 2DT \\ \frac{A}{A'} &= \frac{C}{C'} = \frac{D}{D'} = 2.303R \end{aligned}$$

$$\text{Temperature of maximum ionization, } T_{\max} = \left( \frac{A'}{D'} \right)^{\frac{1}{2}} = \left( \frac{A}{D} \right)^{\frac{1}{2}}$$

$$\text{pK at } T = T_{\max}, \text{pK}_{\max} = 2\sqrt{A'D'} - C'$$

Amino Acid	A'	C'	D'	A	C	D
Glycine pK <sub>1</sub> .....	1300.53	5.5277	.011792	5949.98	25.2894	.053949
Glycine pK <sub>2</sub> .....	3183.69	3.5246	.0060223	14474.01	16.1252	.0412772
<i>dl</i> -Alanine pK <sub>1</sub> .....	1383.06	6.3639	.0136615	6327.55	29.1151	.062502
<i>dl</i> -Alanine pK <sub>2</sub> .....	2941.55	1.8171	.0060945	13457.71	8.3133	.027882
<i>dl</i> - $\alpha$ -Amino- <i>n</i> -Butyric pK <sub>1</sub> .....	1174.74	5.3735	.012487	5374.48	24.5840	.057129
<i>dl</i> - $\alpha$ -Amino- <i>n</i> -Butyric pK <sub>2</sub> .....	2879.31	1.6446	.0060945	13172.96	7.5242	.027882
<i>dl</i> - $\alpha$ -Amino- <i>n</i> -Valeric pK <sub>1</sub> .....	1222.02	5.5238	.012553	5590.79	25.2716	.057430
<i>dl</i> - $\alpha$ -Amino- <i>n</i> -Valeric pK <sub>2</sub> .....	2618.57	-1.1669	.0028693	11980.06	-7.7635	.013127
<i>dl</i> -Norleucine pK <sub>1</sub> .....	1193.30	5.2850	.012130	5459.40	24.179	.055495
<i>dl</i> -Norleucine pK <sub>2</sub> .....	2851.89	1.2891	.005218	13047.52	5.8977	.023872
$\alpha$ -Amino-isobutyric pK <sub>1</sub> .....	1344.95	6.3053	.013924	6153.20	28.8470	.063703
$\alpha$ -Amino-isobutyric pK <sub>2</sub> .....	3010.95	1.5404	.005520	13775.55	7.0474	.025254
<i>dl</i> -Valine pK <sub>1</sub> .....	1245.31	6.0251	.013868	5697.34	27.565	.063447
<i>dl</i> -Valine pK <sub>2</sub> .....	2776.46	1.1033	.005056	12702.42	5.0477	.023131
<i>dl</i> -Leucine pK <sub>1</sub> .....	1283.60	6.0027	.013505	5872.52	27.4626	.061786
<i>dl</i> -Leucine pK <sub>2</sub> .....	2819.38	1.2306	.005127	12698.76	5.6713	.023456
<i>dl</i> -Isoleucine pK <sub>1</sub> .....	1298.09	6.1967	.013959	5938.81	28.3502	.063863
<i>dl</i> -Isoleucine pK <sub>2</sub> .....	2933.52	2.0479	.006578	13420.97	9.3693	.030094

values for the amino acids are not quite so accurate as this; but Smith, Taylor and Smith<sup>8</sup> conclude that "the results in the acid solutions are probably consistent to within  $\pm 0.005$  pK, the results in the alkaline solutions being perhaps less accurate."



TABLE 2A. THERMODYNAMIC FUNCTIONS FOR AMINO ACIDS AT 25° CALCULATED FROM THE EQUATIONS GIVEN IN TABLE 2

Data of B. B. Owen, *J. Am. Chem. Soc.*, **56**, 24 (1934) (glycine); and of P. K. Smith, A. C. Taylor and E. R. B. Smith, *J. Biol. Chem.*, **122**, 109 (1937) (other amino acids).

Amino Acid	pK	$\Delta F^\circ$	$\Delta H^\circ$	$\Delta S^\circ$	$\Delta C_p^\circ$
Glycine pK <sub>1</sub> .....	2.350	3205	1156	-6.9	-32.0
Glycine pK <sub>2</sub> .....	9.778	13336	10806	-8.5	-24.6
Alanine pK <sub>1</sub> .....	2.348	3203	773	-8.2	-37.2
Alanine pK <sub>2</sub> .....	9.867	13453	10980	-8.3	-16.6
$\alpha$ -Amino- <i>n</i> -Butyric pK <sub>1</sub> .....	2.284	3123	298	-9.5	-34.0
$\alpha$ -Amino- <i>n</i> -Butyric pK <sub>2</sub> .....	9.831	13408	10695	-9.1	-16.6
$\alpha$ -Amino- <i>n</i> -Valeric pK <sub>1</sub> .....	2.318	3161	487	-9.0	-34.2
$\alpha$ -Amino- <i>n</i> -Valeric pK <sub>2</sub> .....	9.806	13375	10834	-8.6	-7.8
Norleucine pK <sub>1</sub> .....	2.334	3183	528	-8.9	-33.0
Norleucine pK <sub>2</sub> .....	9.833	13410	10926	-8.3	-14.2
$\alpha$ -Aminoisobutyric pK <sub>1</sub> .....	2.357	3215	492	-9.1	-38.0
$\alpha$ -Aminoisobutyric pK <sub>2</sub> .....	10.206	13919	11531	-8.0	-15.0
Valine pK <sub>1</sub> .....	2.286	3118	59	-10.3	-37.7
Valine pK <sub>2</sub> .....	9.718	13254	10647	-8.8	-13.8
Leucine pK <sub>1</sub> .....	2.329	3176	382	-8.4	-36.8
Leucine pK <sub>2</sub> .....	9.747	13293	10814	-8.3	-13.3
Isoleucine pK <sub>1</sub> .....	2.319	3163	264	-9.7	-38.0
Isoleucine pK <sub>2</sub> .....	9.754	13303	10747	-8.6	-17.9

$\Delta F^\circ$  and  $\Delta H^\circ$  in calories per mole;  $\Delta S^\circ$  and  $\Delta C_p^\circ$  in cal deg<sup>-1</sup> mole<sup>-1</sup>

By accurate calorimetric measurement, J. M. Sturtevant [*J. Am. Chem. Soc.*, **63**, 88 (1941)] finds for glycine  $\Delta H_1 = 930$  cal/mole and  $\Delta H_2 = 10700$  cal/mole. Sturtevant's value  $\Delta H_B = 2771$  and the value of H. S. Harned and B. B. Owen [*J. Am. Chem. Soc.*, **25**, 31 (1939)] of  $\Delta H^\circ = 13480$  cal/mole for the reaction  $H_2O \rightleftharpoons H^+ + OH^-$ . Sturtevant [*J. Am. Chem. Soc.*, **64**, 762 (1942)], with further improved technique, has obtained for *dl*-alanine  $\Delta H_1 = 615$  and  $\Delta H_2 = 10960$  cal/mole.

The values in this table differ slightly from those previously reported by Smith, Taylor and Smith, because of the difference in the form of the equation used for the computations (Table 2).

TABLE 3. pK VALUES FOR TRIVALENT AMINO ACIDS AT 25°, FROM CELLS WITHOUT LIQUID JUNCTION

Amino Acid	pK <sub>1</sub>	pK <sub>2</sub>	pK <sub>3</sub>
Aspartic acid.....	2.01 ( $\alpha$ -COOH)	3.895 ( $\beta$ -COOH)	9.842 ( $-\text{NH}_3^+$ )
Arginine.....	1.807 ( $-\text{COOH}$ )	9.01 ( $-\text{NH}_3^+$ )	not determined
Ornithine.....	1.705 ( $-\text{COOH}$ )	8.690 ( $\alpha$ - $\text{NH}_3^+$ )	10.755 ( $\beta$ - $\text{NH}_3^+$ )

Values from A. C. Batchelder and C. L. A. Schmidt, *J. Physical Chem.*, **44**, 893 (1940).

All the amino acids studied by these authors were of the monoamino-monocarboxylic type.<sup>9</sup> The values given in Table 2a reveal that the pK

values are very similar indeed for all the substances studied. The extreme variation of  $pK_1$  values at  $25^\circ$  is 0.07 (between 2.284 for  $\alpha$ -amino-*n*-butyric acid and valine, and 2.357 for  $\alpha$ -amino-isobutyric acid). On the other hand, the  $pK_2$  values show a much more marked variation, the extreme values at  $25^\circ$  lying at 9.72, for valine, and 10.20 for  $\alpha$ -aminoisobutyric acid. If the latter substance be omitted, however, all the  $pK_2$  values at  $25^\circ$  lie in the range  $9.79 \pm 0.08$ .  $\alpha$ -Amino-isobutyric acid is the only substance yet studied having two methyl groups on the  $\alpha$ -carbon atom; its peculiar behavior may be connected with this structure. Otherwise there is no obvious relation between  $pK$  and structure; a fact which is not surprising, since all these molecules differ only with respect to the number of methylene groups attached to the  $\alpha$ -carbon atom.

The variation with temperature of all these  $pK$  values may be described in terms of an equation developed by Harned and Embree<sup>10</sup>. In a modified form, this is

$$pK = pK_{\max.} + \text{const.} (T_{\max.} - T)^2 \quad (4)$$

where  $T_{\max.}$  is the temperature of maximum ionization and  $pK_{\max.}$  is the  $pK$  value at that temperature. The constant in 4 has the value  $5 \times 10^{-5}$  for the  $pK_1$  values of the amino acids, and  $10 \times 10^{-5}$  for the  $pK_2$  values. Values of  $T_{\max.}$  and  $pK_{\max.}$  for the various amino acids were listed by Smith, Taylor and Smith<sup>8</sup>.

The data in Tables 2 and 2a, however, are described in terms of a different equation, recently employed by Harned and Robinson<sup>10a</sup>.

$$\Delta F^\circ = -RT \ln K = 2.303 RT pK = A - CT + DT^2 \quad (5)$$

$A$ ,  $C$  and  $D$  are empirical constants. Equation 5 gives a distinctly better fit to the experimental data than does 4, for both fatty acids and amino acids.

The heats of ionization,  $\Delta H_1^\circ$ , and  $\Delta H_2^\circ$ , are determined from the temperature coefficients of the ionization constants by the familiar equation of van't Hoff<sup>11</sup>:

$$\Delta H^\circ = RT^2 \frac{\partial \ln K}{\partial T} = -2.303 RT^2 \frac{\partial pK}{\partial T} \quad (6)$$

<sup>10</sup> Harned, H. S., and Embree, H. D., *J. Am. Chem. Soc.*, **56**, 1050 (1934). For a critical discussion of this equation, and others related to it, see K. Fitzer, *J. Am. Chem. Soc.*, **59**, 2365 (1937); R. W. Gurney, *J. Chem. Physics*, **6**, 499 (1938); F. C. Baughman, *ibid.*, **7**, 951 (1939); D. H. Everett and W. F. K. Wynne-Jones, *Trans Faraday Soc.*, **35**, 1359 (1939).

<sup>10a</sup> Harned, H. S., and Robinson, R. A., *Trans. Faraday Soc.*, **36**, 973 (1940).

<sup>11</sup>  $\Delta H_1$  is the heat absorbed, at constant pressure and temperature, in the reaction,

From 5 and 6 we obtain

$$\Delta H^\circ = A - DT^2 \quad (5a)$$

and equations for other thermodynamic functions in these solutions are readily obtained (Table 2). The temperature of maximum ionization (at which  $\partial pK/\partial T = 0$ ) is given by equation 5 as

$$T_{\max.} = (A/D)^{1/2} \quad (5b)$$

and it follows from equations 5a and 5b that  $\Delta H^\circ = 0$  at  $T_{\max.}$ . It is apparent from the values of  $A$  and  $D$  in Table 2, that  $T_{\max.}$  is much greater for  $pK_2$  of the amino acids than for  $pK_1$ . For  $pK_1$  this characteristic

TABLE 4. CHARACTERISTIC HEATS OF IONIZATION FOR VARIOUS GROUPS  
 $\Delta H_i$  IN CALORIES PER MOLE

Substance	Temp.	$\Delta H_i$	Substance	Temp.	$\Delta H_i$
1. Carboxyl Groups $R \cdot \text{COOH} \rightleftharpoons R \cdot \text{COO}^- + \text{H}^+$					
Formic acid	25°	-13	Maleic acid ( $K_1$ )	25°	+210
Acetic acid	25°	-112	Maleic acid ( $K_2$ )	25°	+1280
Propionic acid	25°	-168	Fumaric acid ( $K_1$ )	25°	+900
<i>n</i> -Butyric acid	25°	-691	Fumaric acid ( $K_2$ )	25°	-980
Benzoic acid	13°	+495	<i>d</i> -Tartaric acid ( $K_1$ )	20°	+800
Succinic acid ( $K_1$ )	25°	+710	<i>d</i> -Tartaric acid ( $K_2$ )	20°	+2000
Succinic acid ( $K_2$ )	25°	-340			
2. Phenolic Groups $R \cdot \text{OH} \rightleftharpoons \text{RO}^- + \text{H}^+$					
Phenol	20°	+6100			
Resorcin	20°	+7000			
Hydroquinone	18°	+5900			
3. Ammonium Groups $\text{RNH}_3^+ \rightleftharpoons \text{R} \cdot \text{NH}_2 + \text{H}^+$					
Ammonium	25°	+12800	Anilinium	20°	+7500
Methylammonium	20°	+13000	Piperidinium	20°	+12800
Ethylammonium	20°	+13300	Pyridinium	20°	+5100
Dimethylammonium	20°	+11600			
Trimethylammonium	20°	+8500			

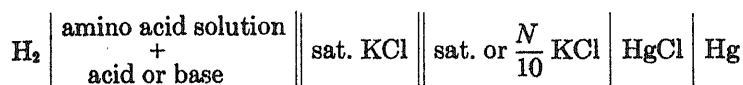
Data from Landolt-Börnstein and International Critical Tables.

temperature is only slightly above room temperature (about 330°K); for  $pK_2$  it is about 300° higher. (Whether  $K_2$  would actually pass through a maximum at the high temperature specified by 5b is of course unknown; the constants in equation 5 are fitted to experimental data obtained between 0° and 60°, and there is no theoretical reason to suppose that they would describe the data adequately at much higher temperatures). From this it follows that  $\Delta H_2^\circ$ , at ordinary temperatures, is far greater than  $\Delta H_1^\circ$ . At 25°, all the  $\Delta H_1^\circ$  values lie between 0 and 1200 cal; the  $\Delta H_2^\circ$  values between 10,700 and 11,100 cal. These values are highly characteristic of the ionization of carboxyl and of ammonium groups respec-

similar earlier data) by Kolthoff<sup>12</sup>, Ebert<sup>13</sup> and Weber<sup>14</sup>. Data on heats of ionization, indeed, often yield decisive evidence as to the nature of an ionizing group in a more complex molecule, as will be illustrated in the later discussion.

### Dissociation Constants from Cells with Liquid Junction

The great majority of dissociation constants hitherto measured for amino acids, and all measurements yet made on peptides, were obtained from hydrogen electrodes involving cells with liquid junctions, generally of the type



The potential of such a cell is *primarily* determined by the hydrogen ion activity (or concentration) in the amino acid solution; but there is inevitably also a potential difference—generally small, but significant—at the junction between the amino acid and the KCl solution. The exact evaluation of this liquid junction potential is a formidable problem<sup>15</sup>, which is far from being completely solved. Furthermore, the great majority of the measurements hitherto made were at a fixed finite ionic strength, without any attempt to extrapolate them to infinite dilution. Thus the constants determined are “apparent dissociation constants”  $\text{pK}'$ . If A is an acid and B is its conjugate base, then the apparent acid dissociation constant is

$$\text{pK}' = \text{pH} + \log \frac{C_A}{C_B} = \text{pK} + \log \frac{\gamma_B}{\gamma_A} \quad (7)$$

where the  $\gamma$ 's denote the activity coefficients, which depend on the charge type and the chemical structure of A and B<sup>16</sup>, and vary with the ionic strength and the dielectric constant of the solvent.

In spite of the uncertainties and limitations of measurements involving liquid junction potentials, these data (Table 5) are of major importance. They give values which in general agree within 0.1 in pH with the more accurate values from cells without liquid junction (Table 2); and they have been extended to a vast variety of molecules, unstudied by any other method. Thus they furnish the chief basis for determining the state of

<sup>12</sup> Kolthoff, I. M., *Rec. Trav. Chim. Pays-Bas*, 44, 68 (1925).

<sup>13</sup> Ebert, L., *Z. Physik. Chem.*, 121, 385 (1926).

<sup>14</sup> Weber, H. H., *Biochem. Z.*, 218, 1 (1930).

<sup>15</sup> Taylor, P. B., *J. Phys. Chem.*, 31, 1478 (1927); Clark, W. M., “*The Determination of Hydrogen Ions*,” 3rd edition, Baltimore, 1928, especially Chapter XXIII. See Chapter 3 of this book.

<sup>16</sup> Taylor, P. B., *J. Phys. Chem.*, 31, 1478 (1927).

TABLE 5. APPARENT DISSOCIATION CONSTANTS OF AMINO ACIDS, PEPTIDES AND RELATED SUBSTANCES DETERMINED FROM EMF MEASUREMENTS ON CELLS WITH LIQUID JUNCTION AT 25°

A. Substances Containing One Amino and One Carboxyl Group

	Reference	pK <sub>1</sub> '(COOH)	pK <sub>2</sub> '(NH <sub>3</sub> <sup>+</sup> )	pI=pM
<i>1. Amino acids</i>				
Glycine.....	5	2.34	9.60	5.97
Alanine.....	5	2.34	9.69	6.00
α-Amino-n-butyric acid.....	5	2.55	9.60	6.08
Oxy-α-aminobutyric acid.....	5	2.71	9.62	6.16
Valine.....	14	2.32	9.62	5.96
α-Amino-n-valeric acid.....	5	2.36	9.72	6.04
Leucine.....	14	2.36	9.60	5.98
Isoleucine.....	14	2.36	9.68	6.02
Norleucine.....	14	2.39	9.76	6.08
Serine.....	14	2.21	9.15	5.68
Proline.....	*	1.99	10.60	8.30
Oxyproline.....	*	1.92	9.73	5.83
Phenylalanine.....	*	1.83	9.13	5.48
Tryptophane.....	18	2.38	9.39	5.89
Methionine.....	6	2.28	9.21	5.74
Isoserine.....	6	2.78	9.27	6.02
Hydroxyvaline.....	4	2.61	9.71	6.15
Taurine.....	1	1.5(SO <sub>3</sub> H)	8.74	5.12
β-Alanine.....	18a	3.60	10.19	6.90
γ-Amino-n-valeric acid.....	18a	4.02	10.40	7.21
δ-Amino-n-valeric acid.....	16	4.270**	10.768**	7.518
ε-Amino-n-caproic acid.....	5a	4.43	10.75	7.59
ω-Amino-n-dodecanoic acid.....	16	4.648	—	—
<i>2. Amino acid amides</i>				
Glycine amide.....	22	—	7.93	—
Glutamine (γ-amide).....	13	2.17	9.13	5.65
Asparagine (β-amide).....	3	2.02	8.80	5.41
β-Hydroxyasparagine.....	3	2.12	8.26	5.19
Isoglutamine (α-amide).....	13	3.81	7.88	5.85
Isosparagine (α-amide).....	13	2.97	8.02	5.50
α-Hydroxyasparagine.....	3	2.31	7.17	4.74
<i>3. Peptides</i>				
Glycylglycine.....	*	3.06	8.13	5.60
	10	3.12	8.17	5.65
	16	3.083**	8.265**	5.674
Glycylalanine.....	*	3.15	8.25	5.70
Glycylleucine.....	*	3.18	8.29	5.73
Glycylvaline.....	*	3.17	8.25	5.71
Alanylalanine.....	*	3.17	8.42	5.79
Alanylglycine.....	*	3.11	8.18	5.64
α-Aminobutyryl-α-aminobutyric acid.....	19 (at 20°)	3.04	8.39	5.72
Glycylproline.....	13	2.84	8.53	5.69
Alanylproline.....	13	3.04	8.38	5.71
Alanyldiglycine.....	19 (at 20°)	3.21	8.15	5.68
Glycylalanylalanylglycine.....	*	3.30	7.93	5.62
Leucyl-octaglycylglycine.....	19 (at 20°)	(2.27)	(7.84)	(5.03)
Phenylalanylglycine.....	10	3.10	7.71	5.41
Triglycine.....	*	3.26	7.91	5.59
	*	3.05	7.75	5.40

TABLE 5—Continued

## 4. Amino acids and peptides containing methylated amino groups

Substance	Reference	pK <sub>1</sub> ' (COOH)	pK <sub>2</sub> ' (Amino)	pI = pM
Sarcosine.....	12	2.23	10.01	6.12
Sarcosylglycine.....	12	3.10	8.51	5.80
Glycyl-sarcosine.....	12	2.83	8.54	5.68
Sarcosylsarcosine.....	12	2.86	9.10	5.98
N-dimethylglycine.....	11c	1.94	9.86	5.90
Betaine.....	20	1.84	Very large	—

## B. Amino Acids and Peptides with One Amino, One Carboxyl and One or More Phenolic or Sulfhydryl Groups

Substance	Reference	pK <sub>1</sub> ' (COOH)	pK <sub>2</sub> '	pK <sub>3</sub> '	pI
L-Tyrosine.....	21	2.20	9.11 (NH <sub>3</sub> <sup>+</sup> )	10.07 (OH)	5.66
Diiodo-L-tyrosine.....	21	2.12	6.48 (OH)	7.82 (NH <sub>3</sub> <sup>+</sup> )	4.29
Dibromo-L-tyrosine.....	21	2.17	6.45 (OH)	7.60 (NH <sub>3</sub> <sup>+</sup> )	4.30
Dichloro-L-tyrosine.....	21	2.12	6.47 (OH)	7.62 (NH <sub>3</sub> <sup>+</sup> )	4.29
3,4-dihydroxyphenylalanine.....	14	2.36	8.68 (NH <sub>3</sub> <sup>+</sup> )	{ 9.88 (OH) 11.68 (OH)	5.51
Glycyltyrosine.....	8	2.98	8.40 (NH <sub>3</sub> <sup>+</sup> )	10.40 (OH)	5.60
Tyrosyltyrosine.....	8	3.52	7.68 (NH <sub>3</sub> <sup>+</sup> )	{ 9.80 (OH) 10.26 (OH)	5.60
Cysteine.....	*(at 30°)	1.96	{ 8.18 (NH <sub>3</sub> <sup>+</sup> ) 8.33** (NH <sub>3</sub> <sup>+</sup> )	{ 10.28 (SH) 10.78** (SH)	{ 5.07 5.02
Cysteinylcysteine.....	†	2.65	7.27 (NH <sub>3</sub> <sup>+</sup> )	{ 9.35 (SH) 10.85 (SH)	4.96

\* For aspartyltyrosine and tyrosylarginine, see below

## C. Substances Containing Two or More Carboxylic Groups and One Basic Group

Substance	Reference	pK <sub>1</sub> ' (COOH)	pK <sub>2</sub> ' (COOH)	pK <sub>3</sub> '	pK <sub>4</sub> '	pI
Aspartic acid.....	14	1.88	3.65	9.60 (NH <sub>3</sub> <sup>+</sup> )	—	2.77
Glutamic acid.....	14	2.19	4.25	9.67 (NH <sub>3</sub> <sup>+</sup> )	—	3.22
	15	2.155**	4.324**	9.960** (NH <sub>3</sub> <sup>+</sup> )	—	3.24
β-Hydroxyglutamic acid.....	14	2.33	4.24	9.56 (NH <sub>3</sub> <sup>+</sup> )	—	3.29
Glycylaspartic acid.....	*	2.81	4.45	8.60 (NH <sub>3</sub> <sup>+</sup> )	—	3.63
Aspartylglycine.....	*	2.10	4.53	9.07 (NH <sub>3</sub> <sup>+</sup> )	—	3.31
Aspartylaspartic acid.....	7	2.70	3.40	4.70 (COOH)	8.26 (NH <sub>3</sub> <sup>+</sup> )	3.04
Glutaminylglycine.....	13	3.15	—	7.52 (NH <sub>3</sub> <sup>+</sup> )	—	—
Glutaminylglutamic acid.....	13	3.14	4.38	7.62 (NH <sub>3</sub> <sup>+</sup> )	—	—
Aspartyltyrosine.....	10	2.13	3.57	8.92 (NH <sub>3</sub> <sup>+</sup> )	10.23 (OH)	2.85
α-Aminotricarballylic acid.....	11	2.10	3.60	4.60 (COOH)	9.82 (NH <sub>3</sub> <sup>+</sup> )	2.87
Glycyl-α-amino-tricarballic acid.....	11	2.70	4.10	5.35 (COOH)	8.32 (NH <sub>3</sub> <sup>+</sup> )	3.41
Glutathione.....	†	2.12	3.53	8.66 (NH <sub>3</sub> <sup>+</sup> )	9.12 (SH)	2.83

## D. Substances Containing One Carboxyl Group and More than One Basic Group

Substance	Reference	pK <sub>1</sub> ' (COOH)	pK <sub>2</sub> '	pK <sub>3</sub> '	pK <sub>4</sub> '	pI
Histidine.....	18	1.82	6.00 (Im)	9.17 (NH <sub>3</sub> <sup>+</sup> )	—	7.59
Arginine.....	*	2.17	9.04 (NH <sub>3</sub> <sup>+</sup> )	12.48 (Guan.)	—	10.76
Ornithine.....	17	1.94	8.65 (NH <sub>3</sub> <sup>+</sup> )	10.76 (NH <sub>3</sub> <sup>+</sup> )	—	9.70
Lysine.....	*	2.18	8.95 (αNH <sub>3</sub> <sup>+</sup> )	10.53 (εNH <sub>3</sub> <sup>+</sup> )	—	9.74
α,β-Diaminopropionic acid.....	9	1.33	6.80 (αNH <sub>3</sub> <sup>+</sup> )	9.80 (βNH <sub>3</sub> <sup>+</sup> )	—	8.20
Histidylhistidine.....	7	2.25	5.60 (Im)	6.80 (NH <sub>3</sub> <sup>+</sup> )	7.80 (NH <sub>3</sub> <sup>+</sup> )	7.30
Histidylglycine.....	10	2.40	5.80 (Im)	7.82 (NH <sub>3</sub> <sup>+</sup> )	—	6.81
1-Methyl histidine.....	23	1.69	6.48 (Im)	8.85 (NH <sub>3</sub> <sup>+</sup> )	—	7.67
Carnosine.....	23	2.64	6.83 (Im)	9.51 (NH <sub>3</sub> <sup>+</sup> )	—	8.17
Anserine.....	23	2.64	7.04 (Im)	9.49 (NH <sub>3</sub> <sup>+</sup> )	—	8.27
Lyxyl-lysine.....	10	1.95	8.17 (NH <sub>3</sub> <sup>+</sup> )	9.45 (NH <sub>3</sub> <sup>+</sup> )	10.63 (NH <sub>3</sub> <sup>+</sup> )	10.04

TABLE 5—Continued  
E. Substances Containing Two Carboxylic and Two Amino Groups (Tetrapoles)

Substance	Reference	pK <sub>1</sub> ' (COOH)	pK <sub>2</sub> ' (COOH)	pK <sub>3</sub> ' (NH <sub>3</sub> <sup>+</sup> )	pK <sub>4</sub> ' (NH <sub>3</sub> <sup>+</sup> )	pI = pM
Cystine.....	16a	1.65	2.26	7.85	9.85	5.05
	2a (at 30°)	<1.00	1.7	7.48	9.02	4.60
	2	1.04**	2.05**	8.00**	10.25**	5.03
Lysylglutamic acid.....	10	2.93	4.47	7.75	10.50	6.10
Diglycylcystine.....	11a (at 35°)	2.71	2.71	7.94	7.94	5.33
Cystinylidiglycine.....	11a (at 35°)	3.12	3.12	6.36	6.95	4.74
Cystinylidiglycine.....	11a (at 35°)	3.21	3.21	6.01	6.87	4.61
α-Aspartylhistidine.....	11b (at 38°)	2.45	3.02	6.82 (Im)	7.98	4.92
β-Aspartylhistidine.....	11b (at 38°)	1.93	2.95	6.93 (Im)	8.72	4.94

pK<sub>3</sub>' in the aspartylhistidine corresponds to the imidazole group

The pK values as given in Table 5 denote what H. S. Simms [*J. Am. Chem. Soc.*, 48, 1239 (1926)] has denoted the "titration constants" (denoted in his terminology by pG) rather than the traditional "dissociation constants". Where the distinction is important the discussion of Simms and the papers from which our pK values are taken should be consulted. See also the discussion of titration constant in Chapter 20 of this book. pI denotes pH of isoelectric point, pM denotes pH of maximum charge.

† Hastings *et al.* (unpublished).

‡ Birch and Harris [*Biochem. J.*, 24, 564 (1930)] give pK<sub>1</sub> = 2.18, pK<sub>2</sub> = 9.09, pK<sub>3</sub> = 13.2 at 23° for arginine; for histidine they give pK<sub>1</sub> = 1.78, pK<sub>2</sub> = 5.97, pK<sub>3</sub> = 8.97.

\* Values marked with an asterisk are taken from the compilation by E. J. Cohn, *Ergebn. d. Physiol.*, 32, 781 (1931). Many of these references are also given by Kirk and Schmidt, *Univ. of Calif. Publication*, 7, 57 (1929).

\*\* Denotes pK values calculated by extrapolation of observed measurements to zero ionic strength.

Supplementary References, in addition to those cited in the *Ergebnisse d. Physiol.*, 33, 781 (1931).

- (1) Andrews and Schmidt, *J. Biol. Chem.*, 73, 651 (1927).
- (2) Borsook, Ellis and Huffman, *J. Biol. Chem.*, 117, 281 (1937).
- (2a) Cannan and Knight, *Biochem. J.*, 21, 1384 (1927).
- (3) Chibnall and Cannan, *Biochem. J.*, 24, 945 (1930).
- (4) Czarnetsky and Schmidt, *J. Biol. Chem.*, 72, 453 (1931).
- (5) Czarnetsky and Schmidt, *Z. physiol. Chem.*, 204, 129 (1931).
- (5a) Edsall and Blanchard, *J. Am. Chem. Soc.*, 55, 2337 (1933).
- (6) Emerson, Kirk and Schmidt, *J. Biol. Chem.*, 92, 449 (1931).
- (7) Greenstein, *J. Biol. Chem.*, 93, 479 (1931).
- (8) Greenstein, *J. Biol. Chem.*, 95, 465 (1932).
- (9) Greenstein, *J. Biol. Chem.*, 96, 499 (1932).
- (10) Greenstein, *J. Biol. Chem.*, 101, 603 (1933).
- (11) Greenstein and Joseph, *J. Biol. Chem.*, 110, 619 (1935).
- (11a) Greenstein, Klemperer and Wyman, *J. Biol. Chem.*, 129, 681 (1939).
- (11b) Greenstein and Klemperer, *J. Biol. Chem.*, 128, 245 (1939).
- (11c) Johnston, *Proc. Roy. Soc. London A* 78, 101 (1906-07).
- (12) Levene, Simms and Pfaltz, *J. Biol. Chem.*, 61, 445 (1924).
- (13) Melville and Richardson, *Biochem. J.*, 29, 187 (1935).
- (14) Miyamoto and Schmidt, *J. Biol. Chem.*, 90, 165 (1931).
- (15) Neuberger, *Biochem. J.*, 30, 2085 (1936).
- (16) Neuberger, *Proc. Roy. Soc. London A* 158, 68 (1937).
- (16a) Sano, *Biochem. Z.*, 168, 14 (1926).
- (17) Schmidt, Kirk and Schmidt, *J. Biol. Chem.*, 81, 249 (1929).
- (18) Schmidt, Appleman and Kirk, *J. Biol. Chem.*, 81, 723; 85, 137 (1929).
- (19) Tillmans, Hirsch and Strache, *Biochem. Z.*, 199, 399 (1928).
- (20) Weider, *Ber.*, 68, 263 (1935).
- (21) Winnek and Schmidt, *J. Gen. Physiol.*, 18, 889 (1935).
- (22) Zief and Edsall, *J. Am. Chem. Soc.*, 59, 2245 (1937).
- (23) Deutsch and Eggleton, *Biochem. J.*, 32, 209 (1938).

ionization of a given molecule at a given pH level, and for the correlation of dissociation constant and structure (Chapter 5)<sup>17</sup>.

<sup>17</sup> Small differences in pK' among the values given in Table 5 should not generally be considered significant. Thus alanine and α-amino-n-butyric acid are recorded in this table as differing by 0.21 in pK<sub>1</sub>; but the true pK<sub>1</sub> values (Table 2) differ only by 0.07. Larger differences in reported pK' values, however, are generally

The probable nature of the acidic group involved in each dissociation constant is indicated in Table 5. For substances containing only amino and carboxylic groups, the assignment can be made unequivocally on the basis of the discussion already given. For substances containing phenolic, sulfhydryl, imidazole, and guanidine groups, the assignment depends on a

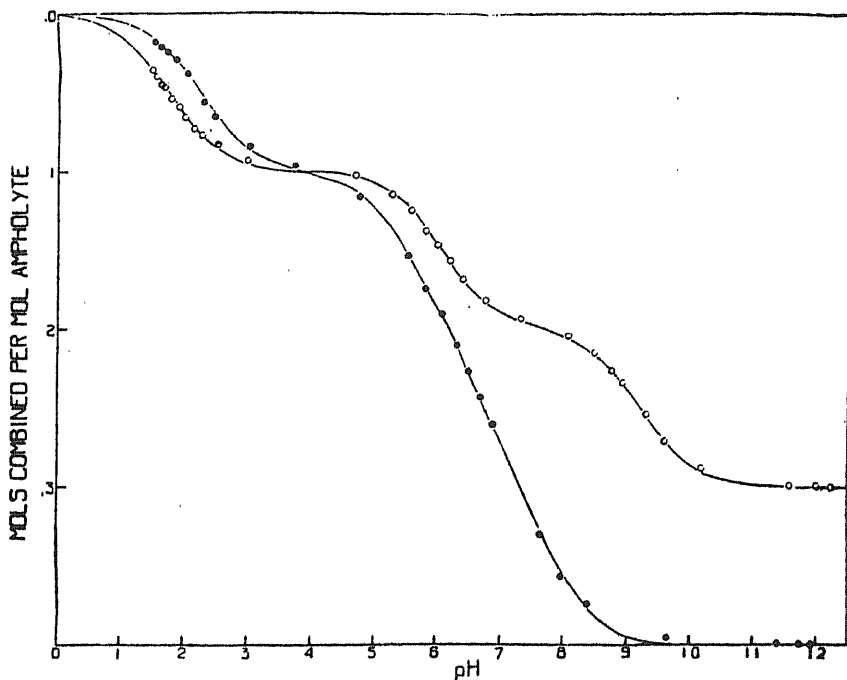


Figure 1. Curves constructed on the basis of the constants: histidine,  $pG_1 = 1.77$ ,  $pG_2' = 6.10$ ,  $pG_3' = 9.18$ ; histidyl-histidine,  $pG_1' = 2.25$ ,  $pG_2' = 5.60$ ,  $pG_3' = 6.80$ ,  $pG_4' = 7.80$ . ● indicates histidyl-histidine; O, histidine. From Greenstein, J. P., *J. Biol. Chem.*, **93**, 479 (1931).

knowledge of the behavior of these groups in simpler compounds. A discussion of this and many other related points is given in Chapter 5.

The titration curves of the monoamino-monocarboxylic amino acids and peptides always show two widely separated regions of marked buffering power, and a broad intermediate zone in which the curve is nearly flat. As the number of ionizing groups is increased, the curves spread out over a wider range; introduction of imidazole groups in the histidine peptides gives rise to buffering power in the neutral pH range; so that



give titration curves qualitatively very similar to those of proteins. (Figs. 1 and 2.) This problem is further discussed in Chapter 20.

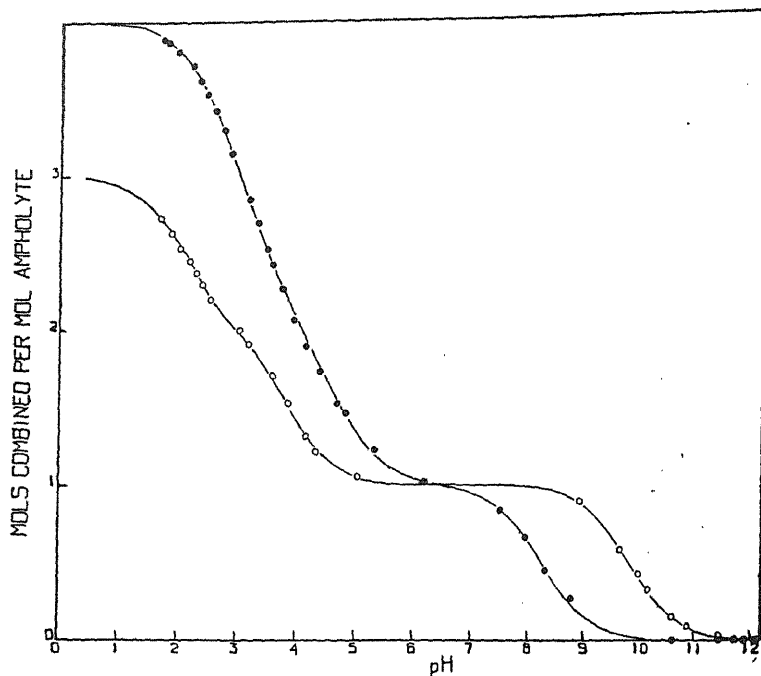


Figure 2. Curves constructed on the basis of the constants:  $G_1' = 2.10$ ,  $pG_2' = 3.86$ ,  $pG_3' = 9.82$ ;  $G_1' = 2.70$ ,  $pG_2' = 4.40$ ,  $pG_3' = 4.70$ ,  $pG_4' = 8.26$ .  $\bullet$  From Greenstein, J. P., *J. Biol. Chem.*, 93, 17.

#### Apparent Heats of Ionization of Amino Acids and Peptides

Apparent dissociation constants at two or more different temperatures have been determined for a number of amino acids and peptides. From the temperature coefficient of  $pK'$ , the "apparent heat of ionization"  $\Delta H'$  may be calculated by the van't Hoff equation, if it be assumed that the heat of ionization is independent of the temperature, over the range of measurement (Table 6).

These apparent heats of ionization agree well in order of magnitude with the much more exact values given in Table 2. The values for the carboxyl group range between +2100 and -1300 calories per mole; for the amino group between +10,000 and +13,300—except for the amino group in histidine and the dihydroxyphenyl group.

TABLE 6. APPARENT HEATS OF IONIZATION OF AMINO-ACIDS AND PEPTIDES  
 $\Delta H'$  (IN CALORIES PER MOLE)

Substance	Ionizing Group			Ammonium	Guanidine
	Carboxyl	Imidazole	Phenolic OH		
Glycine*	0	—	—	10,700	—
$\alpha$ -Alanine*	-450	—	—	11,600	—
$\alpha$ -Amino- <i>n</i> -butyric acid <sup>6</sup>	1200	—	—	13,100	—
Oxy- $\alpha$ -amino- <i>n</i> -butyric acid <sup>5</sup>	1600	—	—	12,100	—
Valine <sup>6</sup>	-300	—	—	10,600	—
Oxyvaline <sup>6</sup>	-1300	—	—	10,700	—
Glycylglycine* <sup>10</sup>	-600	—	—	10,000	—
Glycyl-alanine*	-600	—	—	11,100	—
Phenylalanyl-glycine <sup>10</sup>	680	—	—	10,000	—
Diiodo- <i>l</i> -tyrosine <sup>21</sup>	980	—	810	8,790	—
Dibromo- <i>l</i> -tyrosine <sup>21</sup>	1700	—	860	9,120	—
Dichloro- <i>l</i> -tyrosine <sup>21</sup>	1140	—	1420	8,830	—
Histidine*	1200	6900	—	9,400	—
Histidyl-glycine <sup>10</sup>	300	7500	—	10,800	—
Lysine*	300	—	—	{ (pK <sub>2</sub> ) 11,600 (pK <sub>3</sub> ) 11,350 (pK <sub>2</sub> ) 12,700 (pK <sub>3</sub> ) 11,350 (pK <sub>4</sub> ) 13,300	—
Lysyl-lysine <sup>10</sup>	2000	—	—	{ (pK <sub>2</sub> ) 11,200 (pK <sub>3</sub> ) 10,500 (pK <sub>4</sub> ) 11,950	—
Arginine*	1000	—	—	11,200	12,400
Phenylalanyl-tyrosine <sup>10</sup>	450	—	—	10,150	11,950
Tyrosyl-tyrosine	0	—	6000	10,500	13,000
Aspartic acid*	{ (pK <sub>1</sub> ) 1600 (pK <sub>2</sub> ) 2100	—	—	11,200	—
Aspartyl-tyrosine <sup>10</sup>	{ (pK <sub>1</sub> ) 750 (pK <sub>2</sub> ) 0	—	6200	10,160	—
Glutamic acid*	{ (pK <sub>1</sub> ) 1900 (pK <sub>2</sub> ) 1040	—	—	11,200	—
Lysylglutamic acid <sup>10</sup>	{ (pK <sub>1</sub> ) 750 (pK <sub>2</sub> ) 0	—	—	{ (pK <sub>3</sub> ) 10,500 (pK <sub>4</sub> ) 11,950	—

The reference numbers are the same as in Table 5. The values marked \* are taken from the compilation by E. J. Cohn, *Ergebnisse d. Physiol.*, **33**, 781 (1931).

The above values are calculated from the temperature coefficient of the pK' values by the relation:

$$-\Delta H' = \frac{\Delta(R \ln K)}{\Delta(1/T)}$$

If pK' measurements are made at 0° and at 25° (as for most of the substances recorded here), this becomes

$$-\Delta H' = 14,940(pK'_{25} - pK'_{0}).$$

If the measurements were made at 25° and at 40° (as was the case for the dihalogenated tyrosines), the formula becomes:

$$-\Delta H' = 28,490(pK'_{40} - pK'_{25}).$$

The values given for glycylglycine are those of Branch and Miyamoto (Cohn, *Ergebnisse*). The later measurements of Greenstein (10) give  $\Delta H'(\text{COOH})$  as +680, and  $\Delta H'(\text{NH}_3^+)$  as 10,300 calories per mole.

is near 7000. For the phenolic group in tyrosyl-arginine and aspartyl-

These represent the most marked deviations from regularity among any of the groups found; evidently they are intimately bound up with the effect of halogenation on the benzene ring, and call for further study. The  $\Delta H$  values for the guanidine group (around 12,500) are even higher than for the amino group; on account of the high alkalinity at which the ionization of this group must be studied, these values are probably less accurate than those for the other groups reported.

The accuracy of these values is, in any case, not extremely high; comparison with the exact values of Table 2 shows a discrepancy in  $\Delta H_1$  for alanine of 1250 cal/mole, for instance, and for  $pK_2$  of  $\alpha$ -amino-*n*-butyric acid of 2500 cal/mole. The error in most cases is smaller than this, however, and probably seldom exceeds 1000 cal/mole. In spite of all uncertainties, the data of Table 6 furnish a most valuable guide to the behavior of the molecules concerned, and to the nature of the groups ionizing at various pH values. Similar studies on compounds whose structure is not exactly known should provide important information regarding structure. Wyman's studies on hemoglobin and other studies on proteins (Chapter 20) have already shown the power of the method in analyzing the titration curves of proteins.

### Isoelectric Points of Amino Acids and Peptides

In strongly acid solution, all amino acids and peptides are positively charged and migrate as cations in an electric field; in strongly alkaline solution all are negatively charged and migrate as anions. For any such substance, therefore, some intermediate pH value may be found where the average net charge on the ampholyte molecules is zero and no net movement toward either pole will occur. This pH value is known as the isoelectric point (pI) of the ampholyte.

Following the classical treatment of Michaelis<sup>18</sup> we may readily calculate the isoelectric point of a simple ampholyte with two acidic groups from its dissociation constants. The condition for the isoelectric state is that the average net charge per molecule of ampholyte in the solution should be zero; that is that the number of cations ( $R^+$ ) be equal to the number of anions ( $R^-$ ), assuming that both ions have approximately the same mobility. In terms of the dissociation constants (letting  $R$  represent the concentration of molecules with zero net charge)

$$(R^+) = \frac{(H^+)(R)}{K_1} = (R^-) = \frac{K_2(R)}{(H^+)} \quad (8)$$

Whence

$$(\text{H}_1^+)^2 = K_1 K_2 \quad \text{or} \quad \text{pH}_1 = \frac{\text{p}K_1 + \text{p}K_2}{2} \quad (9)$$

( $\text{H}_1^+$ ) denoting the hydrogen ion concentration at the isoelectric point.

This treatment is strictly valid only at infinite dilution where the activities are equal to the concentrations. At finite concentrations, however, the same equation is valid if we replace  $K_1$  and  $K_2$  by the apparent constants  $K'_1$  and  $K'_2$ . The values of  $K'_1$  and  $K'_2$ , however, are functions not only of the concentration of the ampholyte itself, but of the ionic strength and of the chemical nature of the solvent medium. The  $K'$  values used, therefore, must be those valid for the particular state of the system under consideration.

Actually, for most simple ampholytes of this type, such as glycine or glyceryl-glycine,  $\text{p}K_1$  and  $\text{p}K_2$  are so far apart that there is not merely an isoelectric point, but a broad zone of pH values in which the ampholyte is practically isoelectric. In glycine, for instance, the concentration of  $[\text{R}^+]$  and  $[\text{R}^-]$  is less than one per cent of that of  $[\text{R}^\pm]$  at all pH values between 4.3 and 7.7.

The same form of treatment may be readily extended to more complex ampholytes. As the condition for the isoelectric state, we again assume that the average net charge per molecule of ampholyte shall be zero; that is, that the sum of all the positive charges on the ampholyte molecules shall equal the sum of all the negative charges.

For an ampholyte which can acquire a maximum positive charge of  $m$  units in strongly acid solution, and a maximum negative charge of  $n$  units in strongly alkaline solution, the condition for the isoelectric state is thus:

$$m(\text{R}_{m+}) + (m-1)(\text{R}_{(m-1)+}) + \cdots + 2(\text{R}_{2+}) + (\text{R}_{+}) \\ = (\text{R}_{-}) + 2(\text{R}_{2-}) + \cdots + (n-1)(\text{R}_{(n-1)-}) + n(\text{R}_{n-}) \quad (10)$$

where  $(\text{R}_{m+})$  denotes the concentration of the ions carrying  $m$  positive charges, and so forth. In most amino acids, peptides and proteins,  $m$  is the total number of imidazole, guanidine, and free amino groups in the molecule;  $n$  is the number of free carboxyl, phenolic and sulfhydryl groups.

The dissociation constants of the ampholyte may be written as

$$K_1 = \frac{(\text{H}^+)(\text{R}_{(m-1)+})}{(\text{R}_{m+})}; \cdots; K_{m-1} = \frac{(\text{H}^+)(\text{R}_{1+})}{(\text{R}_{2+})}; K_m = \frac{(\text{H}^+)(\text{R})}{(\text{R}_{1+})} \\ K_{m+1} = \frac{(\text{H}^+)(\text{R}_{1-})}{(\text{R})}; K_{m+2} = \frac{(\text{H}^+)(\text{R}_{2-})}{(\text{R}_{1-})}; \cdots; K_{m+n} = \frac{(\text{H}^+)(\text{R}_{n-})}{(\text{R}_{(n-1)-})} \quad (11)$$

In terms of  $(H^+)$  and the  $K$ 's, equation 10 may be rewritten as

$$\begin{aligned} \frac{m(H^+)^m}{K_1 K_2 \dots K_{m-1} K_m} + \frac{(m-1)(H^+)^{m-1}}{K_2 K_3 \dots K_m} + \dots + \frac{2(H^+)^2}{K_{m-1} K_m} + \frac{(H^+)}{K_m} \\ = \frac{K_{m+1}}{(H^+)} + \frac{2K_{m+1} K_{m+2}}{(H^+)^2} + \dots + \frac{nK_{m+1} K_{m+2} K_{m+n}}{(H^+)^n} \end{aligned} \quad (12)$$

This may be rearranged, for convenience, into the form

$$\begin{aligned} \frac{(H^+)}{K_m} \left( 1 + \frac{2(H^+)}{K_{m-1}} + \frac{3(H^+)^2}{K_{m-2} K_{m-1}} + \dots \right) \\ = \frac{K_{m+1}}{(H^+)} \left( 1 + \frac{2K_{m+2}}{(H^+)} + \frac{3K_{m+2} K_{m+3}}{(H^+)^2} + \dots \right) \end{aligned} \quad (13)$$

If  $K_{m-1} \gg 2(H^+)$  and  $(H^+) \gg 2K_{m+2}$ , all the terms involving  $(H^+)$  in the parentheses are small compared to 1, and as an approximation we may write for the isoelectric point<sup>19</sup>

$$(H^+)^2 = K_m K_{m+1} \quad \text{or} \quad pH_I = \frac{pK_m + pK_{m+1}}{2} \quad (14)$$

These conditions are fulfilled for all amino acids and peptides yet studied. Thus, for cystine ( $m = n = 2$ ), equation 13 becomes

$$\frac{(H^+)}{K_2} \left( 1 + \frac{2(H^+)}{K_1} \right) = \frac{K_3}{(H^+)} \left( 1 + \frac{2K_4}{(H^+)} \right) \quad (15)$$

From the data of Borsook, Ellis and Huffman<sup>20</sup>  $pK_1 = 1.04$ ,  $pK_2 = 2.05$ ,  $pK_3 = 8.00$ ;  $pK_4 = 10.25$ .

The approximate equation 14 gives  $pH_I = (2.05 + 8.00)/2 = 5.03$ ;  $(H^+) = 10^{-5.03}$ .

If we substitute this approximate value of  $(H^+)$  as a trial value in the expressions in brackets in equation 15, and solve to obtain a more exact value for  $(H^+)$ , we obtain

$$(H^+)^2 = K_2 K_3 \left( \frac{1.000012}{1.0002} \right)$$

which is indistinguishable from the approximate value. The same equation (15) and the same conclusions hold also for lysylglutamic acid.

In the case of  $\alpha$ -amino tricarballic acid<sup>21</sup>  $m = 1$ ,  $n = 3$ , and equation 13 becomes:

$$\frac{(H^+)}{K_1} = \frac{K_2}{(H^+)} \left( 1 + \frac{2K_3}{(H^+)} + \frac{3K_3 K_4}{(H^+)^2} \right) \quad (16)$$

<sup>19</sup> S. H. H. S. J. *Am. Chem. Soc.* 65, 1943 (1943). See also Hitchcock, D. I., *J. Biol. Chem.* 114, 111 (1942).

The  $pK$  values are  $pK_1 = 2.10$ ,  $pK_2 = 3.60$ ,  $pK_3 = 4.60$ ,  $pK_4 = 9.82$ . The value of  $(H^+)$  must lie somewhere between  $K_1$  and  $K_2$ ; hence it is apparent at once that the last term on the right hand side is negligible. The approximate equation 14 gives  $pH_I = 2.85$ . Substituting this value for  $(H^+)$  into the parenthesis, and solving again we obtain  $pH_I = 2.843$ ; so even in this rather unfavorable case the correction term is only  $-0.007$ .

For the dibasic acids, histidine, arginine, and lysine,  $m = 2$ ,  $n = 1$ , and equation 13 becomes

$$\frac{(H^+)}{K_2} \left( 1 + \frac{2(H^+)}{K_1} \right) = \frac{K_3}{(H^+)} \quad (17)$$

and it may readily be seen from the  $pK$  values in Table 5 that the approximate equation  $pH_I = \frac{pK_2 + pK_3}{2}$  holds within the limits of accuracy of the experimental measurements.

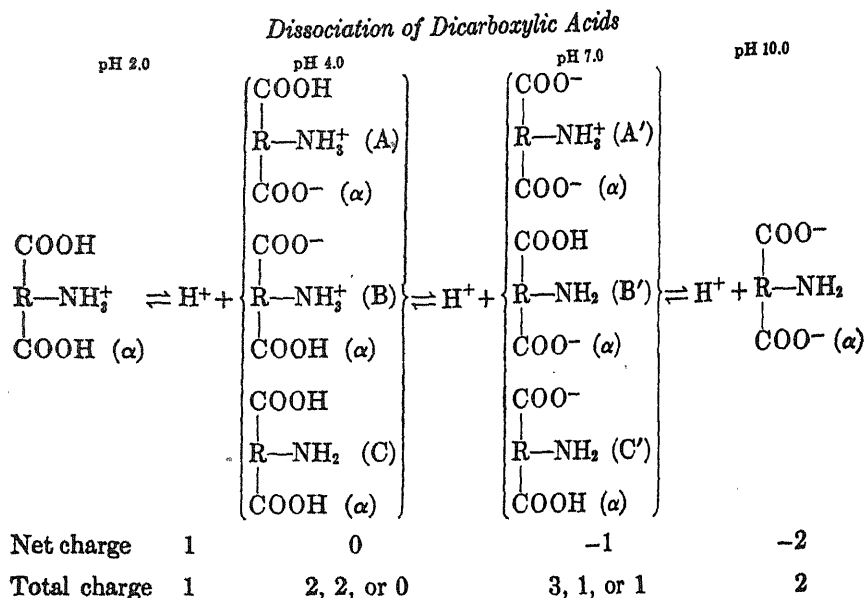
Here again the equations as written are strictly valid only at infinite dilution; at finite concentrations we must substitute for each  $K$  the corresponding appropriate value of  $K'$ .

The only quantities needed in this treatment are the values of  $m$  and  $n$ , and the experimentally determined  $K$  values. It is unnecessary to establish any correlation between the values of  $K$  and the nature of the individual ionizing groups.

#### Net Charge and Total Charge on Ampholyte Molecules: The Point of Maximum Charge

The study of dipolar ions has revealed the necessity of distinguishing between the total charge of a molecule and its net charge<sup>22</sup>. Thus the total charge on a glycine molecule is a maximum ( $= 2$ ) at its isoelectric point, while its net charge (zero) is a minimum at this point. Let us consider now the general case of an ampholyte in strongly acid solution, where it carries its maximum net positive charge. Addition of an equivalent of base will cause removal of a proton from one of the groups present. If the removal occurs from an uncharged acidic group, the total charge increases by one; if from a cationic group, the total charge decreases by one; the net charge decreases by one in either case. Consider for example the successive steps in the ionization of dicarboxylic amino acids.

Of the three different possible isoelectric forms represented below, (A) will predominate, since a proton tends to dissociate more readily from the  $\alpha$ -carboxyl group (strengthened by the positive charge on the ammonium



group) than from the distal carboxyl or the ammonium group. The distal carboxyl dissociates next, giving rise to the form (A'), which therefore is present in much larger amount than are B' and C'. A' is clearly the form of maximum charge; and the dicarboxylic amino acids therefore carry a maximum total charge when the net charge on the molecule is -1.

We may now evaluate approximately the pH of maximum total charge, making the assumption that the forms (C), (B') and (C') above are present in negligible amounts. Then we may write for the three dissociation constants of a dicarboxylic acid

$$K_1 = \frac{(\text{H}^+)(\text{R}^\pm)}{(\text{R}^+)}; \quad K_2 = \frac{(\text{H}^+)(\text{R}_\pm^-)}{(\text{R}^\pm)}; \quad K_3 = \frac{(\text{H}^+)(\text{R}^-)}{(\text{R}_\pm^-)} \quad (18)$$

The symbols are chosen so as to denote the total charge carried by each ionic form. The average total charge,  $Q$ , per molecule of ampholyte is

$$Q = \frac{(\text{R}^+) + 2(\text{R}^\pm) + 3(\text{R}_\pm^-) + 2(\text{R}^-)}{(\text{R}^+) + (\text{R}^\pm) + (\text{R}_\pm^-) + (\text{R}^-)} \quad (19)$$

In terms of  $(\text{H}^+)$  and the  $K$ 's, this becomes

$$Q = \frac{1 + \frac{2K_1}{(\text{H}^+)} + \frac{3K_1K_2}{(\text{H}^+)^2} + \frac{2K_1K_2K_3}{(\text{H}^+)^3}}{\dots} \quad (20)$$

At the point of maximum charge,  $Q$  is a maximum and  $\frac{\partial Q}{\partial (H^+)} = 0$ . On differentiation, it is found that this condition is satisfied by the relation

$$1 + \frac{4K_2}{(H^+)} + \frac{3K_2K_3}{(H^+)^2} + \frac{K_1K_2}{(H^+)^2} = \frac{K_1K_2^2K_3}{(H^+)^4} \quad (21)$$

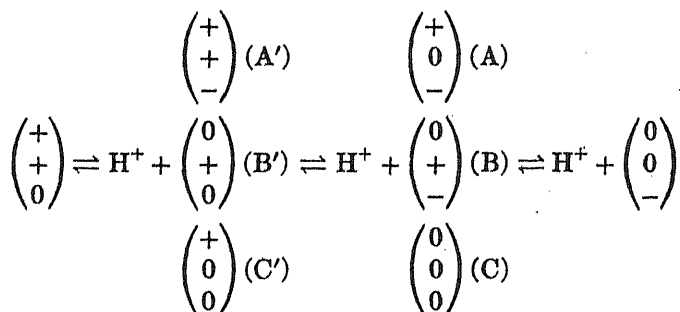
Since  $(H^+)$  at the point of maximum charge must lie between  $K_2$  and  $K_3$ , the terms 1 and  $\frac{3K_2K_3}{(H^+)^2}$  are small in comparison with the others on the lefthand side. Omitting them, and rearranging, we obtain for the point of maximum charge,  $(H_M^+)$ :

$$(H_M^+)^2 = \frac{K_1K_2K_3}{K_1 + 4(H_M^+)} \quad (22)$$

Since in general  $K_1 \gg 4(H_M^+)$  for these substances, this assumes approximately the simple form<sup>23</sup>

$$(H_M^+) = \sqrt{K_2K_3} \quad \text{or} \quad pH_M = \frac{pK_2 + pK_3}{2} \quad (23)$$

An exactly similar treatment may be given for the dibasic amino acids. Here the successive steps in dissociation may be represented by the scheme



where the sets of three figures in parentheses denote, reading *upward*: (1) charge on the carboxyl groups: zero or minus; (2) charge on  $\alpha$ -amino group: zero or plus; (3) charge on  $\epsilon$ -amino, guanidino, or imidazole group: zero or plus. The state of maximum charge corresponds to the predominance of the form  $\begin{pmatrix} + \\ + \\ - \end{pmatrix}$ , which is produced when the isoelectric ampholyte

has combined with one mole of acid. The approximate equation for the point of maximum charge is

$$nK_1 + nK_2$$



The total charge on an ampholyte molecule is important because it largely determines the electrical interaction between the ampholyte and the adjacent solvent molecules. Thus the electrostriction (see Chapter 7) produced by the ampholyte is in general a maximum at the point of maximum charge. Likewise the careful studies of Pertzoff<sup>24</sup> on the optical rotation of aspartic and glutamic acids and their salts have shown that the change of specific rotation with concentration ( $m$ ) is greatest for the acid salts. "The same rule appears to hold for both aspartic and glutamic acids: the slope of the curves giving ( $\alpha$ ) as a function of  $m$  or  $\sqrt{m}$  is directly proportional to the number of charges" (<sup>24</sup>p. 159). Further study may be expected to reveal other situations in which the concepts of total charge and maximum charge are of importance.

These considerations are incomplete in one important respect. In deriving the equations for the point of maximum charge, we have considered only one of the possible isoelectric forms of the ampholyte. Even the simplest amino acids, however, exist in two isoelectric forms—the dipolar ion and the uncharged molecule. More complicated ampholytes also give rise to a variety of dipolar ionic forms. All these forms are in equilibrium with each other, and we shall now set forth a method for evaluating the relative amounts of these different forms in the equilibrium mixture.

#### The Equilibrium Between Dipolar Ions and Uncharged Molecules

We may consider first the simple monoamino-monocarboxylic acids and peptides. Here only two isoelectric forms are involved. These may pass into one another either by intramolecular rearrangement, or by interchange of protons with water molecules and with other acids and bases in the system.

The ratio of their activities may be denoted by  $K_Z$

$$\frac{(R^{\pm})}{(R)} = \frac{(^+H_3N \cdot R \cdot COO^-)}{(H_2N \cdot R \cdot COOH)} = K_Z \quad (25)$$

In an infinitely dilute aqueous solution of the amino acid, which will be taken as the standard state in all that follows,  $K_Z$  is equal to the ratio of the concentrations (or mole fractions) of dipolar ions and uncharged molecules in solution. In any other state, at the same temperature and pressure,  $K_Z$  is by definition the activity ratio of the two forms. The value of this constant is thus an important index of many physico-chemical properties of the system containing the amino acid.

The most satisfactory method available for the determination of  $K_Z$  is one due to Ebert<sup>25</sup>, which involves a comparison of the dissociation

cation ( $R^+$ ) of the amino acid may give off a hydrogen ion from either the carboxyl or the charged amino group; thus two dissociation constants are involved

$$\frac{(H^+)(R^+)}{(R^+)} = K_A \quad (26) \qquad \frac{(H^+)(R)}{(R^+)} = K_B \quad (27)$$

Likewise each of the two isoelectric forms may give off a hydrogen ion to form the amino acid anion

$$\frac{(H^+)(R^-)}{(R^\pm)} = K_C \quad (28) \qquad \frac{(H^+)(R^-)}{(R)} = K_D \quad (29)$$

In terms of these constants, then,

$$K_Z = \frac{(R^\pm)}{(R)} = \frac{K_A}{K_B} = \frac{K_D}{K_C} \quad (30)$$

The dissociation constants, as determined directly by titration, are related as follows to the constants given above

$$K_1 = \frac{(H^+)[(R^\pm) + (R)]}{(R^+)} = K_A + K_B \quad (31a)$$

$$K_2 = \frac{(H^+)(R^-)}{[(R^\pm) + (R)]}; \quad \frac{1}{K_2} = \frac{1}{K_C} + \frac{1}{K_D} \quad (31b)$$

$K_1$  and  $K_2$  can be determined by direct experiment. If we know in addition any one of the four constants  $K_A$ ,  $K_B$ ,  $K_C$  and  $K_D$ , the values of the other three are uniquely determined by equations 30, 31a and 31b. Of these constants,  $K_B$  can be determined as the dissociation constant of the methyl or ethyl ester of the amino acid in question, if we assume that the  $-\text{COOH}$  group is identical with the  $-\text{COOCH}_3$  or  $-\text{COOC}_2\text{H}_5$  group in its effect on the dissociation of a neighboring group. That this assumption is a close approximation to the truth has been shown by Wegscheider.<sup>26, 27</sup> We assume, then,

$$K_B = \frac{(H_2N \cdot R \cdot \text{COOH})(H^+)}{(^+H_3N \cdot R \cdot \text{COOH})} = \frac{(H_2N \cdot R \cdot \text{COOC}_2\text{H}_5)(H^+)}{(^+H_3N \cdot R \cdot \text{COOC}_2\text{H}_5)} = K_E \quad (32)$$

In contrast to the earlier equations, which are exact, equation 32 is only approximate. Assuming  $K_B = K_E$ , however, we may derive from Equations 26–31 inclusive the following relations:

$$K_A = K_1 - K_E \quad K_D = \frac{K_1 K_2}{K_E} \quad K_C = \frac{K_1 K_2}{K_1 - K_E} \quad (33)$$

<sup>26</sup> Wegscheider, R., *Monatsh.*, 16, 153 (1895); 23, 287 (1902).

$$K_1K_2 = K_AK_C = K_BK_D = K_EK_D \quad (34)$$

and for  $K_Z$

$$K_Z = \frac{K_A}{K_B} = \frac{K_A}{K_E} = \frac{K_1}{K_E} - 1 \quad (35)$$

Hence, if  $K_1$  is very much larger than  $K_E$  (which proves to be the case for all the aliphatic amino acids) we have, as a close approximation

$$K_Z = K_1/K_E \quad (36) \quad \text{and} \quad \log K_Z = \text{p}K_E - \text{p}K_1 \quad (37)$$

and likewise

$$K_D = K_2K_Z \quad (38) \quad \text{and} \quad \text{p}K_D = \text{p}K_2 - \log K_Z \quad (39)$$

For the aminobenzoic acids, however, the conditions for this approximation do not hold, and equation 35 must be employed.

The data at present available on amino acids which have been studied by this method are set forth in Table 7.

It is clear that the values of  $K_Z$  are very great for all the aliphatic amino acids, ranging from about 40,000 for glycylglycine to nearly 1,000,000 for  $\epsilon$ -aminocaproic acid. In spite of some uncertainty as to the accuracy of the assumptions involved in this evaluation of  $K_Z$ , they can leave no doubt that the dipolar ion is overwhelmingly predominant in all these solutions.

The aminobenzoic acids present a marked contrast to the aliphatic amino acids. The values of  $K_Z$  for all of them are not far from unity. The *meta* acid, however, is seen to exist predominantly as a dipolar ion in water; the *ortho* and *para* acids predominantly as uncharged molecules. This characteristic difference between the three isomeric acids is also shown clearly by measurements of dielectric constant (Chapter 6) and of apparent molal volume (Chapter 7). The underlying factors determining this difference between the *meta* acids on the one hand and the *ortho* and *para* on the other are discussed in Chapter 5.

The values of  $\text{p}K_D$  (about 4.3 for the  $\alpha$ -amino acids as compared to 4.8 for the unsubstituted fatty acids) represent the effect of an *uncharged* amino group on the dissociation of a neighboring carboxyl. Evidently the amino group produces a moderate increase in the acidity of the carboxyl (in the aliphatic amino acids but not in the aminobenzoic acids); a result in accord with the views of Vorländer<sup>28</sup> and Adams<sup>3</sup>. The calculated values of  $\text{p}K_D$  are somewhat uncertain, owing to the nature of the theory by which they are derived; but the uncertainty is hardly great enough to invalidate this conclusion. (See also Chapter 5, Table 4.)

Several other systematic relationships emerge from the table. In the

identical dissociation constants ( $pK_E = 7.7 \pm 0.1$ ), regardless of the length of the attached carbon chain, a result to be expected from known data on other homologous series. The enormous effect of the  $-\text{COOC}_2\text{H}_5$  group (and presumably therefore of the  $-\text{COOH}$  group) in decreasing the basicity of the amino group is clearly revealed. A primary amine is approximately a thousand times as strong a base as the ester of an  $\alpha$ -amino

TABLE 7. LOGARITHMIC DISSOCIATION CONSTANTS OF AMINO ACIDS AND THEIR ESTERS IN WATER AT 25°

Substance	$pK_1$	$pK_2$	$pK_{\text{Ester}}$	$\log K_a = \log \left( \frac{K_1}{K_E} - 1 \right)$	$pK_D$
1. Glycine.....	2.31	9.72	7.73 (ethyl) 7.66 (methyl)	5.42	4.30
2. $\alpha$ -Alanine.....	2.39	9.72	7.80	5.41	4.31
3. $\alpha$ -Amino- <i>n</i> -Butyric Acid.....	2.55	9.60	7.71	5.16	4.44
4. Leucine.....	2.34	9.64	7.63	5.29	4.35
5. $\beta$ -alanine.....	3.60	10.19	9.13	5.53	4.66
6. $\gamma$ -Amino- <i>n</i> -Butyric Acid.....	4.23	10.43	9.71	5.48	4.95
7. $\delta$ -Amino- <i>n</i> -Valeric Acid.....	4.27	10.77	10.15	5.88	4.89
8. $\epsilon$ -Amino- <i>n</i> -Caproic Acid.....	4.43	10.75	10.37	5.94	4.81
9. Glycylglycine.....	3.14	8.07	7.75	4.61	3.46
10. <i>o</i> -Aminobenzoic Acid.....	2.03	4.98	2.09	-0.70	4.92
11. <i>m</i> -Aminobenzoic Acid.....	3.04	4.79	3.56	+0.36	4.27
12. <i>p</i> -Amino benzoic Acid.....	2.32	4.92	2.38	-0.87	4.86
13. <i>N</i> -dimethyl- <i>o</i> -Amino-benzoic Acid.....	1.4	8.42	5.55	+4.1	4.3
<b>pK Values of Dicarboxylic Acids and Their Esters</b>					
		$\alpha \text{ COOH}$	Distal COOH	$-\text{NH}_2^+$	
14. Aspartic Acid.....	2.08	—	3.87	9.85	
15. Diethyl Aspartate.....	—	—	—	6.5	
16. Glutamic Acid.....	2.155	—	4.324	9.960	
17. $\alpha$ -Ethyl Hydrogen Glutamate.....	—	—	3.846	7.838	
18. $\gamma$ -Ethyl Hydrogen Glutamate.....	2.148	—	—	9.19	
19. Ethyl Glutamate.....	—	—	—	7.035	

References: Nos. 1, 2, 3, 4, 5, 8, 9, 14, 15: Edsall and Blanchard, *J. Am. Chem. Soc.*, 55, 2337 (1933). Nos. 6 and 7: Neuberger, *Proc. Roy. Soc. London (A)*, 158, 68 (1937) where new data on glycylglycine and its ethyl ester are also given. Nos. 10 and 11: Cumming, *Proc. Roy. Soc. London (A)*, 78, 138 (1906-07). No. 12: Johnston, *Proc. Roy. Soc. London (A)*, 78, 101 (1906-07). No. 13: Edsall and Wyman, *J. Am. Chem. Soc.*, 57, 1964 (1935). Nos. 16, 17, 18, and 19: Neuberger, *Biochem. J.*, 30, 2085 (1936).

The papers by Cumming and Johnston, here cited, contain a large amount of further data on methylated aminobenzoic acids and their esters.

acid This effect is so great that even the charged  $\text{COO}^-$  group in the amino acids reduces ten-fold the basicity of a neighboring amino group, although the electrostatic effect of the negative charge on the group would in itself work powerfully in the opposite direction (Chapter 5).

As the distance between the amino group and the  $-\text{COOC}_2\text{H}_5$  group

approximately represented by an equation first developed by MacInnes<sup>29</sup> which is (in our notation)

$$pK' = pK_{\infty} - (S/d) \quad (40)$$

in which  $pK_{\infty}$  is the dissociation constant, when the substituent is at an infinite distance along the carbon chain<sup>30</sup>,  $S$  is a constant characteristic of the substituent, and  $d$  (the "distance") is taken as the number of carbon atoms separating the dissociating group from the substituent. For the amino acid esters, equation 40 becomes

$$pK' = 10.8 - (3.0/d) \quad (41)$$

MacInnes has shown this equation to hold for a variety of substituents, and Schmidt<sup>31</sup> has shown that it applies approximately to both dissociating groups in the amino acids. In Fig. 3 we have plotted  $pK_1$ ,  $pK_2$ , and  $pK_E$  as a function of  $1/d$  for the amino acids and their esters, utilizing Schmidt's data for the amino acids (to which we have added our own determinations for  $\epsilon$ -aminocaproic acid). The figure shows also the values of the pH of the isoelectric point,  $pI$  (calculated from the equation  $pI = (pK_1 + pK_2)/2$ ) and the values of  $\log K_Z$  and  $pK_D$ , as a function of the reciprocal of the distance between the groups.

It is evident that the linear relationship is valid only as a first approximation. Particularly the values of  $pK_2$  and  $pK_E$ , when  $1/d$  is small, deviate considerably from the straight line. The reasons for this deviation, when the substituent and the dissociating group are close together, have been largely revealed by recent work of Kirkwood and Westheimer, which is discussed in Chapter 5.

MacInnes suggested that the term  $S/d$  in equation 40 might represent the effect upon the potential energy of the molecule produced by the approach of charged groups to the distance  $d$ , and that this potential energy should be closely related to the change in the free energy of ionization ( $RT \ln K$ ) of the molecule. If we picture the dissociating group as a charged sphere, at whose surface the electric potential is  $V_0$  in the absence of an external field, then charges  $e_1, e_2, \dots$  at distances  $d_1, d_2, \dots$  will alter the value of the potential to

$$V = V_0 + \frac{e_1}{d_1} + \frac{e_2}{d_2} + \dots = V_0 + \sum \frac{e}{d} \quad (42)$$

if we neglect the distortion of the relative positions of the charges due to their mutual induction. Similarly, for a dissociating compound containing several substituent groups (which we treat as point charges) we may write

<sup>29</sup> MacInnes, D. A., *J. Am. Chem. Soc.*, **50**, 2587 (1928). See also Greenstein, J. P., *ibid.*, **58**, 1314 (1936).  
<sup>30</sup>  $pK'$  is close to, but not in general identical with, the  $pK$  value ( $pK_0$ ) for the corresponding unsubsti-

$$pK' = pK_{\infty} - \frac{S_1}{d_1} - \frac{S_2}{d_2} - \dots \quad (43)$$

Thus for aspartic acid diethyl ester we should calculate, from equations 40, 41, and 43

$$pK' = 10.8 - \frac{3.0}{1} - \frac{3.0}{2} = 6.3 \quad (44)$$

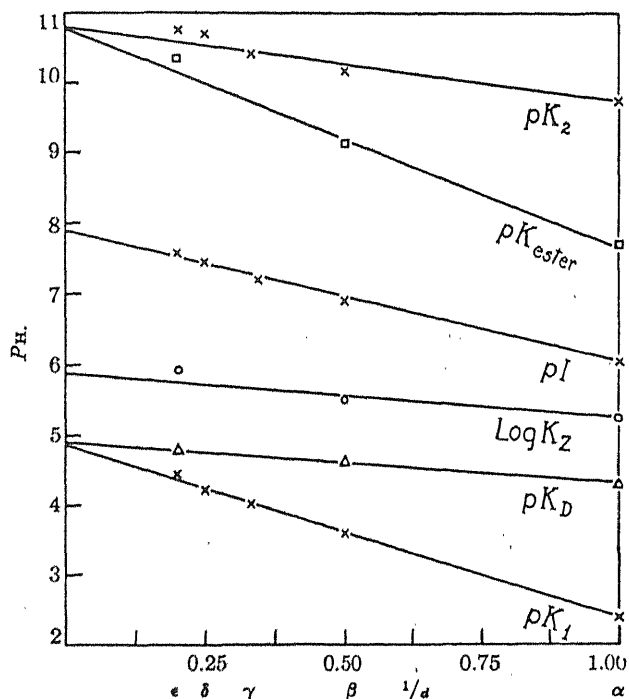
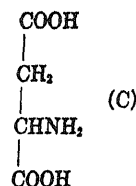
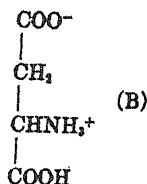
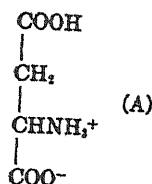


Figure 3. Dissociation constants and related constants for amino acids and their esters, as a function of the reciprocal distance between the polar groups. From Edsall, J. T., and Blanchard, M. H., *J. Am. Chem. Soc.*, 55, 2337 (1933).

whereas the observed  $pK'$  value is 6.5. The agreement is as good as could be expected, considering the uncertainty of the assumptions made and the neglect of the distortion effect. Equation 43 is at any rate empirically useful in making approximate estimates of dissociation constants, when exact values are unknown. For instance, we may make use of it in estimating the relative numbers of the three isoelectric forms of aspartic



A and B are dipolar ions, B having a much higher dipole moment than A; C is an uncharged molecule. Any one of the three forms may arise when a hydrogen ion is given off from the cation  $\text{HOOC} \cdot \text{CH}_2 \cdot \text{CH}(\text{NH}_3^+) \cdot \text{COOH}(\text{D})$ . Three dissociation constants are therefore involved

$$\frac{(\text{H}^+)(\text{A})}{(\text{D})} = K_A \quad (45)$$

$$\frac{(\text{H}^+)(\text{B})}{(\text{D})} = K_B \quad (46)$$

$$\frac{(\text{H}^+)(\text{C})}{(\text{D})} = K_C \quad (47)$$

Hence

$$K_A/(\text{A}) = K_B/(\text{B}) = K_C/(\text{C}) \quad (48)$$

$pK_C$  may be taken as equal to the  $pK$  value of aspartic acid diethyl ester (6.5). For the effect of the  $\text{NH}_3^+$  group on the dissociation of the carboxyl in (A) or (B) we may use Greenstein's<sup>32</sup> equation  $pK_{\text{COOH}} = 4.83 - (2.50/d)$ ; for the effect of the other carboxyl we may use the equation (which holds for the mono esters of the dicarboxylic acids beyond succinic acid)

$$pK_{\text{COOH}} = 4.8 - (0.5/d)$$

Hence, assuming a summation of these effects according to equation (43)

$$pK_A = 4.8 - \frac{2.50}{1} - \frac{0.5}{2} = 2.05$$

$$pK_B = 4.8 - \frac{2.50}{2} - \frac{0.5}{2} = 3.3$$

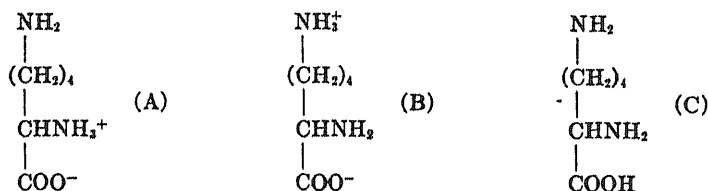
$$pK_C = pK_E = 6.5$$

$pK_A$  is, as it should be, approximately equal to the measured value of  $pK_1$  (2.08). Hence A, B, and C are present in solution approximately in the proportion 28,000 to 1600 to 1. The concentration of the uncharged molecule is negligible; that of the highly polar form B is appreciable; the form A, corresponding to the dipolar ion of the simple amino acids, predominates.

Recently Neuberger<sup>33</sup> has carried out very accurate measurements of the dissociation constants of glutamic acid, diethyl glutamate, and  $\alpha$  and  $\gamma$  monoethyl glutamates. These measurements show that in glutamic acid there are about 48 dipolar ion molecules for every uncharged molecule.

position to every one with the positive charge in the  $\gamma$  position. As in aspartic acid, the concentration of uncharged molecules is negligible. Numerous other interesting relations between dissociation constant and structure have been pointed out in the study by Neuberger.

A similar calculation may be applied to a dibasic amino acid such as lysine which may exist in the three isoelectric forms



Any one of these may lose a hydrogen ion to form the anion (D)  $\text{NH}_2-(\text{CH}_2)_4-\text{CHNH}_2-\text{COO}^-$ . The three dissociation constants involved are then

$$\frac{(\text{H}^+)(\text{D})}{(\text{A})} = K_A \quad (49) \qquad \frac{(\text{H}^+)(\text{D})}{(\text{B})} = K_B \quad (50) \qquad \frac{(\text{H}^+)(\text{D})}{(\text{C})} = K_C \quad (51)$$

Hence

$$K_A(\text{A}) = K_B(\text{B}) = K_C(\text{C}) \quad (52)$$

In estimating  $\text{p}K_A$  and  $\text{p}K_B$  it is probably best to ignore the effect of the distant  $-\text{NH}_2$  group on the dissociation of the charged  $-\text{NH}_3^+$  group, which is probably very small and will be about the same for both the forms A and B. To obtain the value of  $\text{p}K_A$  we may then employ Greenstein's equation:  $\text{p}K_{\text{NH}_3^+} = 10.72 - (0.9/d)$ . Hence

$$\text{p}K_A = 10.72 - \frac{0.9}{1} = 9.8$$

$$\text{p}K_B = 10.72 - \frac{0.9}{5} = 10.54$$

while  $\text{p}K_C$  may be taken as equal to  $\text{p}K_D$  as deduced from equation 39 for the monoamino-monocarboxylic acids:  $\text{p}K_C = 4.3$  ( $\text{p}K_B$ , it may be noted should be approximately equal to the measured value of  $\text{p}K_3$  for lysine (10.53) and this is indeed the case). The ratio of A to B to C in solution is then roughly, from equation 52, 320,000 to 1,800,000 to 1. Here, then, the very highly polar form (B) is predominant. The dielectric constants of lysine solutions should be very high, approaching those of  $\epsilon$ -aminocaproic acid (Chapter 6) and lysine may thereby exert an important

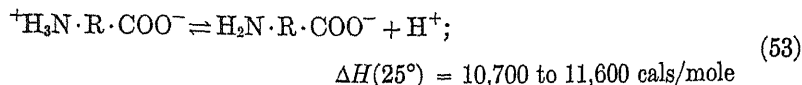


a negative charge on the carboxyl) predominates even more than the corresponding form in lysine. Histidine, on the other hand, is a less polar dipolar ion, similar to the monoamino-monocarboxylic acids.

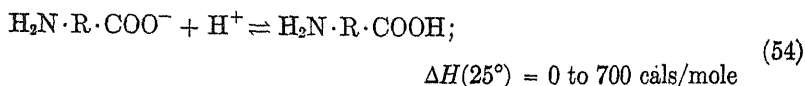
In molecules containing a larger number of acidic and basic groups, it is obvious that many different forms of dipolar ions may exist. In a complicated molecule like a protein, indeed, the possible number is extremely large, although the number of different forms which exist in appreciable quantities is probably relatively small. Further consideration of this important question will be deferred to the chapter dealing with the titration curves of the proteins and their interpretation (Chapter 20).

#### The Effect of Temperature on $K_Z$

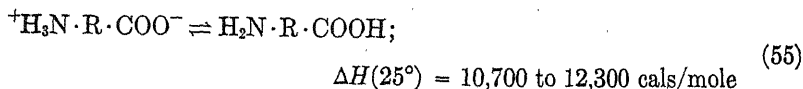
We may estimate this effect by calculating the heat absorbed in the reaction  ${}^+\text{H}_3\text{N}\cdot\text{R}\cdot\text{COO}^- \rightarrow \text{H}_2\text{N}\cdot\text{R}\cdot\text{COOH}$ . From the data on heats of ionization in Table 2, we may write (at 25°)



If we assume that the carboxyl group in  $\text{H}_2\text{N}\cdot\text{R}\cdot\text{COOH}$  is like that in a fatty acid, we may write (from the data in Chapter 5, Table 1).



adding equations 53 and 54



Thus a large amount of heat is absorbed when a dipolar ion is transformed into an uncharged molecule. Hence  $K_Z$  decreases with rise of temperature, being approximately halved for a rise of 10°, according to the equation

$$\frac{\partial \ln K_Z}{\partial T} = - \frac{\Delta H}{RT^2} \quad (56)$$

This temperature effect may be very important in the aminobenzoic acids, where  $K_Z$  is not far from unity; a relatively small shift of temperature might determine whether dipolar ions or uncharged molecules predominate in their solutions. This would have an important effect on the dielectric constant and other properties of the solution.

The temperature coefficient of  $\Delta H$  is given by the equation  $\partial \Delta H / \partial T = \Delta C_p$

$\alpha$ -amino acid, so that  $\Delta H$  increases with increasing temperature. The mechanism of this heat capacity change is discussed in detail in Chapter 7.

### The Relative Concentrations of Dipolar Ions and Uncharged Molecules in Solvents of Lower Dielectric Constant

Highly polar molecules are most soluble in highly polar solvents; slightly polar molecules in less polar or non-polar solvents<sup>34</sup>. The electrical free energy of charged groups increases as the dielectric constant of the solvent is lowered. Since the free energy of the system as a whole tends to a minimum (at constant temperature and pressure) dipolar ions in solvents of low dielectric constant must tend to rearrange themselves into their uncharged isomers. The magnitude of the shift may be evaluated as follows.

### Relative Acidity of Acids of Different Charge Types in Different Media

The preceding discussion has been restricted to acids and their conjugate bases in dilute aqueous solution. The relative strengths of different acids, however, may not be the same if these acids are studied in different solvents. Following Brönsted<sup>2</sup>, however, we may formulate certain general relations which permit a correlation between chemical structure and change in acidity with change in solvent. Consider any acid, A, and its conjugate base, B. The "acidity constant" of this acid in any medium, M, may be denoted by

$$K^* = (H^+) \left( \frac{C_B}{C_A} \right)_M = (H^+) \left( \frac{\gamma_A}{\gamma_B} \right)_M \left( \frac{a_B}{a_A} \right) \quad (57)$$

where the  $C$ 's denote the concentrations of B and A, the  $\gamma$ 's denote activity coefficients, and the  $a$ 's activities, and  $(H^+)$  is the hydrogen ion (proton) activity, which may be taken as equal to the negative antilogarithm of the measured pH<sup>35</sup>.

In a dilute aqueous solution, which will be taken as the standard state, the acidity constant is equal to  $K$  as ordinarily defined,

$$K = (H^+) \left( \frac{C_B}{C_A} \right)_{H_2O} = (H^+) \left( \frac{a_B}{a_A} \right)_{H_2O} \quad (58)$$

since the activities  $a_B$  and  $a_A$  are by definition equal to the concentrations  $C_B$  and  $C_A$ , in water at infinite dilution.  $K$  and  $K^*$  are thus related by the equation (in logarithmic form)

$$pK^* = pK + \log \frac{\gamma_B}{\gamma_A} \quad (59)$$

We have evaluated the activity ratio of dipolar ions to uncharged molecules ( $K_Z$ ) in water from the ratio of  $K_1$ , (amino acid) to  $K_E$  (amino acid ester cation) (equations 35, 36, 37). From equation 57 the same relation yields the concentration ratio of these two forms in any other solvent, if  $K^*$  is substituted for  $K$ . Let a dipolar ion be denoted by  $R^\pm$ , a cation by  $R^+$ , and an uncharged molecule by  $R$ . Then, in any medium,  $M$ ,

$$pK_1^* = pH + \log \left( \frac{C_{R^\pm}}{C_{R^+}} \right)_M \quad (60a)$$

$$pK_E^* = pH + \log \left( \frac{C_{R^\pm}}{C_R} \right)_M \quad (60b)$$

Hence

$$(pK_E^* - pK_1^*) = \log \left( \frac{C_{R^\pm}}{C_R} \right)_M \quad (61)$$

Furthermore, since  $K_Z$  is by definition an *activity* ratio, it is constant for any medium at a given pressure and temperature, thus

$$K_Z = \left( \frac{C_{R^\pm}}{C_R} \right)_{H_2O} = \frac{a_{R^\pm}}{a_R} = \left( \frac{\gamma_{R^\pm}}{\gamma_R} \right)_M \left( \frac{C_{R^\pm}}{C_R} \right)_M \quad (62)$$

or, since  $\log K_Z = (pK_E - pK_1)_{H_2O}$ ,

$$\log \left( \frac{\gamma_{R^\pm}}{\gamma_R} \right)_M = (pK_E - pK_1)_{H_2O} - (pK_E^* - pK_1^*)_M \quad (63)$$

In Table 8, the concentration ratio of dipolar ions to uncharged molecules in 90% alcohol, calculated from equations 61, 62 and 63 is given, along with the activity coefficient ratios.

It is clear that the ratio of dipolar ions to uncharged molecules is much smaller in 90% alcohol than in water; nevertheless the dipolar ion still predominates very greatly. It would be only in media of much lower dielectric constant that the amino acid would exist mainly in the uncharged form<sup>36</sup>.

The relative activity coefficients of amino acids and their uncharged isomers (last column of Table 8) show the profound influence of the very large dipole in increasing the electrical free energy of the molecule in a solvent of low dielectric constant. A similar activity coefficient ratio may be evaluated by determining the relative solubilities of amino acids and their uncharged isomers (such as  $\alpha$ -hydroxyamides) in water and in other solvents, as is shown in detail in Chapters 8 and 9. These entirely independent methods of evaluating the activity coefficient ratio give results which to a first approximation are in very satisfactory agreement.

It is well to examine somewhat further the implications of equations 57 and 58. Consider two different acids,  $A_1$  and  $A_2$ , and their conjugate bases,  $B_1$  and  $B_2$ . If we dissolve (say)  $A_1$  and  $B_2$  in water, they interact according to the equation



the process involving the transfer of a proton from  $A_1$  to  $B_2$ . The equilibrium constant of this reaction may be written, following Brönsted<sup>2b</sup> as:

$$K_{A_1B_2} = \frac{K_1}{K_2} = \frac{a_{A_2} \cdot a_{B_1}}{a_{B_2} \cdot a_{A_1}} = \left( \frac{C_{A_2} \cdot C_{B_1}}{C_{B_2} \cdot C_{A_1}} \right)_{H_2O} \quad (65)$$

TABLE 8. RELATIVE CONCENTRATIONS AND ACTIVITY COEFFICIENTS OF SOME DIPOLAR IONS AND THEIR UNCHARGED ISOMERS IN 90% ALCOHOL

Amino acid	$pK_E^*$	$pK_1^*$	$\log K_2$	$\left( \frac{C_{R\pm}}{C_R} \right)_{alc}$	$\left( \frac{\gamma_{R\pm}}{\gamma_R} \right)_{alc}$
Glycine.....	6.8	3.8	5.42	1000	250
$\alpha$ -Alanine.....	6.8	4.1	5.41	500	500
$\beta$ -Alanine.....	8.2	5.1	5.53	1250	250
Leucine.....	6.8	4.1	5.29	500	400
$\epsilon$ -Aminocaproic acid.....	9.45	6.65	5.94	600	1400

From J. T. Edsall and M. H. Blanchard, *J. Am. Chem. Soc.*, 55, 2352 (1933).

If the reaction 64 takes place in another medium,  $M$ , the protolysis constant, expressed in terms of the concentrations of the reactants, is (see 59):

$$K_{A_1B_2}^* = \frac{K_1^*}{K_2^*} = \left( \frac{C_{A_2} \cdot C_{B_1}}{C_{B_2} \cdot C_{A_1}} \right)_M = \frac{K_1}{K_2} \left( \frac{\gamma_{B_2} \cdot \gamma_{A_1}}{\gamma_{A_2} \cdot \gamma_{B_1}} \right)_M \quad (66)$$

Or, taking logarithms and rearranging terms:

$$(pK_2 - pK_1)_{H_2O} - (pK_2^* - pK_1^*)_M = \log \left( \frac{\gamma_{A_2} \cdot \gamma_{B_1}}{\gamma_{B_2} \cdot \gamma_{A_1}} \right)_M \quad (67)$$

Equation 63 is a special case of 67, in which  $A_1$  is the same as  $A_2$ , being a dibasic acid ( $^+H_2N \cdot R \cdot COOH$ ) which can dissociate to give two different conjugate bases:  $B_1$ , the dipolar ion, and  $B_2$ , its uncharged isomer.

Consider in the general case the nature of the activity coefficient ratio  $\gamma_B/\gamma_A$ , for a given base and its conjugate acid.  $B$  differs from  $A$  primarily by the loss of a proton, and secondarily by any rearrangements in molecular structure brought about by that loss. Therefore we may expect that  $\gamma_B$  and  $\gamma_A$  should be closely related, differing primarily through the change in state of electric charge brought about by the proton transfer. Therefore the ratio  $\gamma_B/\gamma_A$  should be a function primarily of the number and arrange-

Thus for instance let A be a fatty acid,  $R \cdot \text{COOH}$ , and B the conjugate base,  $R \cdot \text{COO}^-$ . The activity coefficient of  $R \cdot \text{COOH}$  in any medium is profoundly affected by the hydrocarbon residue R. Thus acetic acid is very soluble in both water and organic solvents; stearic acid is almost completely insoluble in water, fairly soluble in alcohol, and readily soluble in ether. The two acids, however, have very nearly the same relative strengths in different solvents in which they can be compared. Thus the contribution of R to  $\gamma_B$ , for either acid, must be nearly the same as for  $\gamma_A$ ; the ratio  $\gamma_B/\gamma_A$  is thus determined almost entirely, for either acid, by the loss of a proton by the carboxyl group. Thus we may write, as a good approximation:

$$\frac{\gamma_B}{\gamma_A} = \frac{(\gamma_e)_B}{(\gamma_e)_A} \quad (68)$$

Here  $\gamma_e$  represents the electrostatic contribution to the activity coefficient. Since the  $-\text{COOH}$  group is uncharged, we set  $(\gamma_e)_{\text{COOH}} = 1$  in all solvents. The value of  $(\gamma_e)_{\text{COO}^-}$  may be roughly approximated by treating the group as a charged sphere of radius  $b$ , where  $b$  is of the order of magnitude of 1 to 3 Ångströms. Then, from Born's equation for the electrical free energy of a charged sphere in a medium of dielectric constant  $D$  (Chapter 3, equation 130 for the case in which  $\kappa = 0$ ) we obtain:

$$\ln \gamma_e = 2.303 \log \gamma_e = \frac{\epsilon^2}{2kTb_{\text{COO}^-}} \left( \frac{1}{D} - \frac{1}{D_0} \right) \quad (69)$$

where  $\epsilon$  is the protonic charge,  $D_0$  is the dielectric constant in the standard state (water) and  $D$  is the dielectric constant of the medium, M. The electrical activity coefficient of an ammonium ion,  $R \cdot \text{NH}_3^+$  is approximately given by the same equation, replacing  $b_{\text{COO}^-}$  by  $b_{\text{NH}_3^+}$ .

A similar approximation to the electrical characteristics of a dipolar ion is obtained by treating it as a structure resembling a dumb-bell, composed of two spheres of equal radius,  $b$ , one carrying a charge  $+\epsilon$ , the other a charge  $-\epsilon$ . The spheres are rigidly connected, the distance between their centers being  $R$ <sup>37</sup>. The electrical activity coefficient of such a model is given by the formula<sup>38</sup>, analogous to 69.

$$\ln \gamma_e = 2.303 \log \gamma_e = \frac{\epsilon^2}{kT} \left( \frac{1}{b} - \frac{1}{R} \right) \left( \frac{1}{D} - \frac{1}{D_0} \right) \quad (70)$$

Here  $b$  should be of the order of magnitude of 1 Å, and  $R$  (for an  $\alpha$ -amino acid) of 3 Å.

We may illustrate these considerations with a few of the data of Michaelis

and Mizutani<sup>39</sup>, giving the relative acidities of acids of four different charge types in water and in 90% ethanol. (Table 9.) The pK values as listed by these authors are given; also the relative pK values in each medium, taking pK<sub>1</sub> (glycine) = 0 in each medium as an arbitrary standard, in order to show more closely the marked variation in relative strengths between the two media.

Consider the reaction which may be formulated, following equation 64:

$$\text{CH}_3\text{COOH}(\text{A}_1) + \text{C}_2\text{H}_5\text{NH}_2(\text{B}_2) \rightleftharpoons \text{CH}_3\text{COO}^-(\text{B}_1) + \text{C}_2\text{H}_5\text{NH}_3^+(\text{A}_2) \quad (71)$$

From Table 9, pK<sub>2</sub> - pK<sub>1</sub> = 10.8 - 4.7 = 6.1 in water, while in 90% ethanol pK<sub>2</sub>\* - pK<sub>1</sub>\* = 9.8 - 7.1 = 2.7. Owing to the difference in charge type, the strength of the two acids is much more nearly the same in the medium of lower dielectric constant. Applying equations 67 and 68<sup>39a</sup>

$$\log \left( \frac{\gamma_{\text{A}_2} \cdot \gamma_{\text{B}_1}}{\gamma_{\text{B}_2} \cdot \gamma_{\text{A}_1}} \right) = \log [(\gamma_{\text{NH}_3^+})(\gamma_{\text{COO}^-})]_{\text{Alc.}} = 6.1 - 2.7 = 3.4 \quad (72)$$

TABLE 9. RELATIVE pK' VALUES FOR ACIDS OF DIFFERENT CHARGE TYPE IN WATER AND 90% ETHANOL

Acid	Water		90% Ethanol	
	pK	ΔpK	pK	ΔpK
Glycine cation (pK <sub>1</sub> ).....	2.54	0.00	3.79	0.00
Acetic acid.....	4.70	2.16	7.10	3.31
Glycine (pK <sub>2</sub> ).....	9.81	7.27	9.99	6.20
Ethylammonium ion.....	10.82	8.28	9.78	5.99

Data from L. Michaelis and M. Mizutani, *Z. Physik. Chem.*, 116, 135 (1925).

We may now apply 69, taking  $D_0 = 78.54$  for water at 25°, and  $D = 29.95$  for 90% alcohol, whence  $1/D - 1/D_0 = .0205$ . Taking  $\epsilon = 4.8 \times 10^{-10}$  esu.,  $k = 1.38 \times 10^{-16}$  erg/deg and  $T = 298^\circ$ , we have

$$\log (\gamma_{\text{NH}_3^+} \cdot \gamma_{\text{COO}^-}) = 3.4 = \frac{23.04 \times 10^{-20} \times .0205}{4.606 \times 1.38 \times 10^{-16} \times 298} \left( \frac{1}{b_{\text{NH}_3^+}} + \frac{1}{b_{\text{COO}^-}} \right)$$

$$\text{or} \quad \left( \frac{1}{b_{\text{NH}_3^+}} + \frac{1}{b_{\text{COO}^-}} \right) = 1.36 \times 10^8 \text{ cm}$$

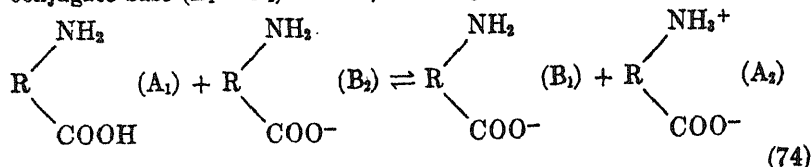
Taking a mean radius,  $b$ , for the two groups, defined by

$$\left( \frac{1}{b_{\text{NH}_3^+}} + \frac{1}{b_{\text{COO}^-}} \right) = \frac{2}{b} \quad (73)$$

we obtain  $b = 1.47 \text{ \AA}$ , which is of the right order of magnitude for these two groups. It is not of course to be expected that equation 69 should give values of  $b$  agreeing exactly with the size of the ionizing groups, as known from molecular structure determinations; for the approximations

made in applying 69 to organic acids and bases deprive it of exact quantitative validity. It is, however, a very valuable guide to the general character of the change of acid-base equilibrium with change in medium in any reaction of this sort.

We have already seen (equation 63 and Table 8) how equation 67 may be employed to evaluate the relative activity coefficients of dipolar ions and their uncharged isomers. A second, independent method of evaluating this activity coefficient ratio may be developed from the data of Table 9. Consider the dipolar ion of an amino acid ( $A_2$ ), and its uncharged isomer ( $A_1$ ), as monobasic acids, each of which may lose a proton to form the same conjugate base ( $B_1 = B_2$ ). Then, following 64, we may write:



The acid,  $\text{H}_2\text{N} \cdot \text{R} \cdot \text{COOH}$  ( $R$ ), and its conjugate base,  $\text{H}_2\text{N} \cdot \text{R} \cdot \text{COO}^-$  ( $R^-$ ), are of the same charge type as  $\text{CH}_3\text{COOH}$  and its conjugate base,  $\text{CH}_3\text{COO}^-$ . Hence from 68,  $\gamma_B/\gamma_A$  should be approximately the same for either pair. Therefore we set

$$(\text{p}K_R - \text{p}K_{\text{HAC}})_{\text{H}_2\text{O}} = (\text{p}K_R^* - \text{p}K_{\text{HAC}})_M \quad (75)$$

Whence, applying 67 and 75 to 74, denoting the dipolar ion by  $R^+$ ,

$$(\text{p}K_{R^+} - \text{p}K_{\text{HAC}})_{\text{H}_2\text{O}} - (\text{p}K_{R^+}^* - \text{p}K_{\text{HAC}})_M = \log \frac{\gamma_{R^+} \gamma_R}{\gamma_{R^-} \gamma_R} = \log \frac{\gamma_{R^+}}{\gamma_R} \quad (76)$$

Substituting numerical values ( $\Delta pK$ ) from Table 9 [where  $\text{p}K_2$  (glycine) =  $\text{p}K_{R^+}$ ]:

$$\log \frac{\gamma_{R^+}}{\gamma_R} = (7.26 - 2.16) - (6.20 - 3.31) = 2.22 \quad (77)$$

while Table 8 gives the somewhat higher value of 2.4. The approximations involved in the derivation of equations 63 and 76 are such that the moderate discrepancy between them is not surprising<sup>40</sup>. Taking  $\log \frac{\gamma_{R^+}}{\gamma_R} = \log \gamma_{\text{dipolar ion}} = 2.3$  as a reasonable average value for 90% ethanol, and substituting in equation 70 for the "dumb-bell" model of a dipolar ion, we get<sup>41</sup>

$$\frac{1}{r} - \frac{1}{R} = 0.46 \times 10^8 \text{ cm}^{-1} \quad (78)$$

If we assume  $R = 3\text{\AA}$ , this gives  $b = 1.27\text{\AA}$ , instead of 1.47 from 72. Here again a simple electrostatic model yields results which are of the right order of magnitude.

Jukes and Schmidt<sup>42</sup> have determined values of  $pK_1^*$  and  $pK_2^*$  in 72% ethanol at 25°, for a number of amino acids, and for glycine ethyl ester. From some of their data, we have calculated values of  $\log \gamma_{R\pm}/\gamma_R$  in this medium for several amino acids (Table 10), by means of both equations 63 and 76. In general, the latter equation gives slightly lower values, but the agreement between the two is reasonably satisfactory. Dibasic and dicarboxylic amino acids were also studied by Jukes and Schmidt.

TABLE 10. RELATIVE ACTIVITY COEFFICIENTS OF DIPOLAR IONS AND THEIR UNCHARGED ISOMERS, FOR AMINO ACIDS IN 72% ETHANOL

Amino Acid	$\log \gamma_{R\pm}/\gamma_R$ (Equation 63)	$\log \gamma_{R\pm}/\gamma_R$ (Equation 76)
Alanine.....	1.73	1.41
Asparagine.....	1.46	1.47
Glycine.....	1.58	1.63
Isoleucine.....	1.87	1.58
Valine.....	1.82	1.60
$\alpha$ -Aminovaleric acid.....	1.78	1.55
$\gamma$ -Aminovaleric acid.....	1.87	1.72
$\delta$ -Aminovaleric acid.....	2.07	1.93
Proline.....	1.58	1.76

Calculated from the data of J. Jukes and C. L. A. Schmidt, *J. Biol. Chem.*, **105**, 359 (1934). Values for  $pK_1^*$  and  $pK_2^*$  (amino acids) were used in applying equation 63, and  $pK_1^*$  (amino acid) as given by Jukes and Schmidt. Values for  $pK_1^*$  and  $pK_2^*$  (amino acids) were used in applying equation 76, along with values of  $pK_1^*$  and  $pK_2^*$  (amino acids).

### The Effect of Varying Ionic Strength on Acid-Base Equilibria Involving Dipolar Ions

Hitherto we have considered only the relative strength of acids and bases in solvents of different dielectric constants. However, the general equations 59, 66, and 67 are immediately applicable to studies on acids and bases carried out at varying ionic strengths in an otherwise constant medium. We shall confine our treatment to the effect of solvents of relatively low ionic strength on the ionization of the same four types of acids already listed in Table 9. We must therefore consider the effect of ionic strength on the activity coefficients of cations, anions, dipolar ions and uncharged molecules. This discussion will be confined to aqueous solutions, although it may readily be generalized to apply to other solvents, from the treatments given in Chapters 3 and 12.

The activity coefficient of an uncharged molecule  $R$  in water is generally given approximately as a function of the ionic strength,  $I/2$ , as

$$\log \gamma_R = K_s I/2 \quad (78)$$



the number of attached hydrocarbon residues increases.  $\log \gamma_{R^+}$  for a dipolar ion is also proportional (at low ionic strengths) to the first power of the ionic strength. It is composed of two terms; a salting out term such as would be characteristic of an uncharged isomer of the dipolar ion, and an electrostatic term of opposite sign, which is a function of the dipole moment and of the arrangement of charges within the molecule (see Chapters 11 and 12).

$$\log \gamma_{R^+} = \log \gamma_s + K_s \frac{I}{2} = (-K_R + K_s) \frac{I}{2} = K'_R \frac{I}{2} \quad (79)$$

where  $K_R$  is always positive, and increases with the dipole moment. Thus if we assume that  $K_s$  is the same for a dipolar ion as for its uncharged isomer, we have

$$\log \frac{\gamma_{R^+}}{\gamma_R} = -K_R \frac{I}{2} \quad (80)$$

Since  $K_R$  is always positive, this means that, in dilute solution, increase of ionic strength increases the concentration ratio of dipolar ions to uncharged molecules.  $\log \gamma$  for an anion or cation in water at low ionic strengths is again made up of a salting out term, depending on the organic residue attached to the charged group, and of an electrical term which (from the Debye-Hückel theory) is proportional to  $\sqrt{I/2}$

$$\log \gamma_{R^+} \cong \log \gamma_{R^-} = \frac{-0.5z^2\sqrt{I/2}}{1 + a\sqrt{I/2}} + K_s I/2 \quad (81)$$

where  $z$  is the valence of the anion or cation ( $z = 1$  in the cases considered here), and  $a$  is proportional to the "collision diameter" of the ions<sup>48</sup>.

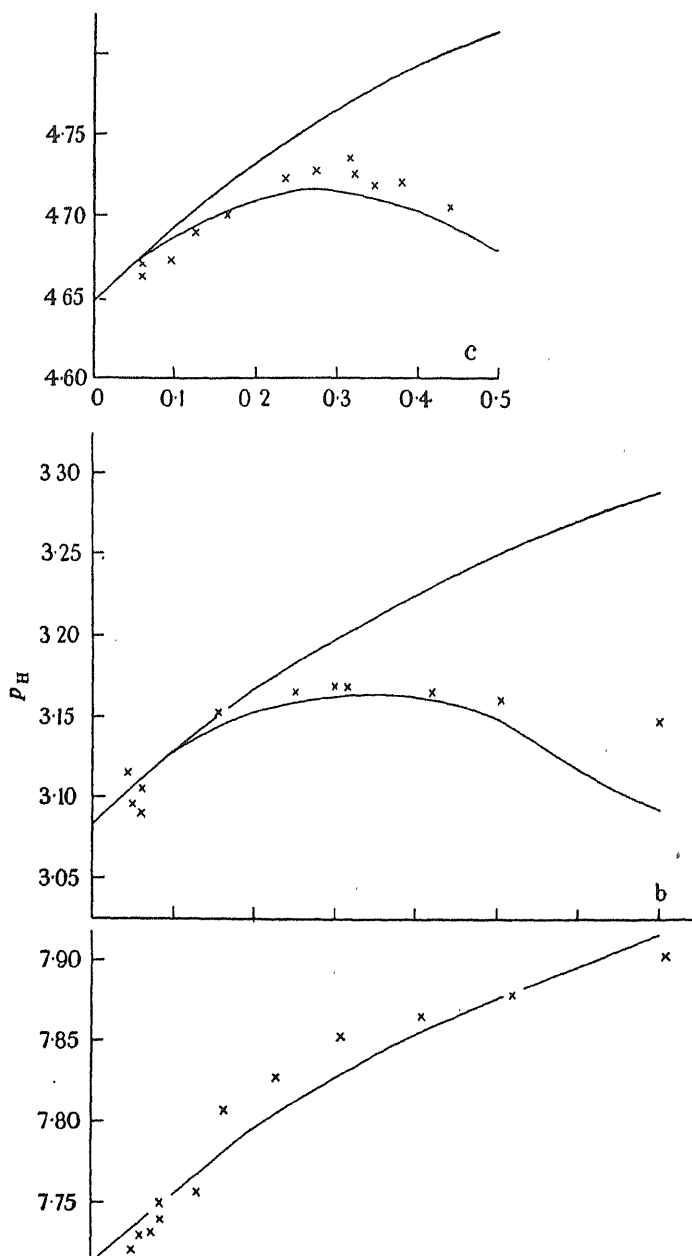
Following the usual convention, we shall write  $pK'$  (instead of  $pK^*$  as in 57, 59, and 60) for the apparent dissociation constant at ionic strength  $I/2$ , and  $pK_0$  for the  $pK$  value in the standard state at zero ionic strength. We may now obtain equations for  $pK'$  in the four types of acids considered in Table 9 by applying equation 59 to each acid and its conjugate base in turn, making use of equations 78, 79, and 81.

Class 1 (exemplified by  $pK_1$  of glycine): A is a cation, B a dipolar ion.

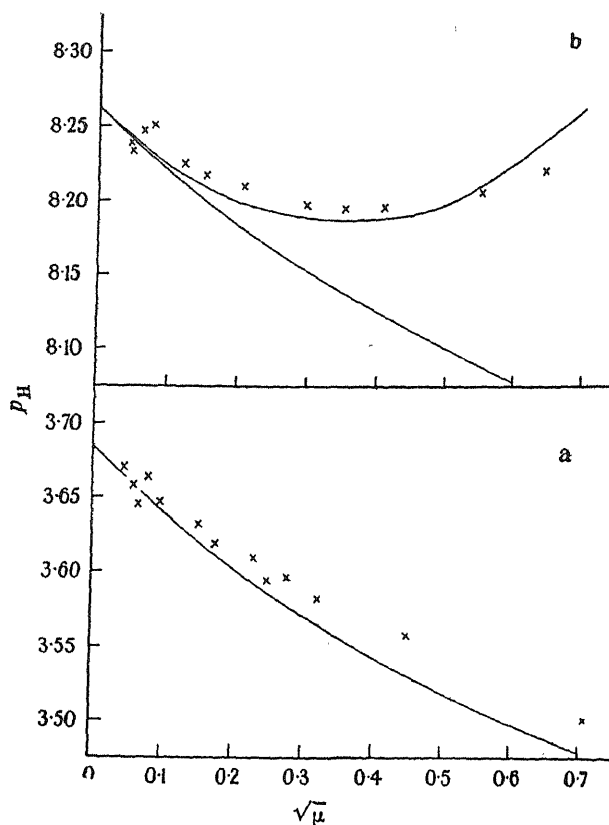
$$pK' = pK_0 + \frac{0.5\sqrt{I/2}}{1 + a\sqrt{I/2}} - K_R \frac{I}{2} \quad (82)$$

Class 2 (exemplified by acetic acid): A is uncharged, B an anion

$$pK' = pK_0 - \frac{0.5\sqrt{I/2}}{1 + a\sqrt{I/2}} \quad (83)$$



la, 4b, and 4c. X Experimental. a, glycylglycine ester; b, glycylglycine pK<sub>1</sub>; c,  $\omega$ -aminododecanoic acid. The curve in and the upper curves in Figures 4b and 4c have been plotted according to equation 85, Chapter 4. The lower curve in s plotted according to equation 82, with  $K_R$  equal to 0.4, and in Figure 4c with  $K_R$  equal to 0.55. From A. Neuburger, *of the Royal Society*, London, A158, 68 (1937).



Figures 5a and 5b.  $\times$  Experimental. a, N-acetyl-glycine; b, glycylglycine  $pK_1$ . The curve in Figure 5a and the lower curve in Figure 5b are plotted according to equation 83, Chapter 4. The upper curve in Figure 5b is plotted according to equation 84 with  $K_R$  equal to 0.4. From A. Neuberger, *Proc. Roy. Soc., London*, A158, 68 (1937).

$$pK' = pK_0 - \frac{0.5\sqrt{I/2}}{1 + a\sqrt{I/2}} + K_R \frac{I}{2} \quad (84)$$

Class 4 (exemplified by the ethylammonium ion): A is a cation, B is uncharged

$$pK' = pK_0 + \frac{0.5\sqrt{I/2}}{1 + a\sqrt{I/2}} \quad (85)$$

In deriving these equations from 78, 79, and 81 we have assumed that  $K_s$  is the same for any acid A as for its conjugate base B. Therefore the

Acids of all four types were studied by Neuberger<sup>44</sup> who derived equations 82–85 inclusive. For simplicity, Neuberger set the “collision diameter” of the ions equal to 3.08 Å in all cases, thus making  $a = 1$  in these equations. His results are shown in Figs. 4 and 5. They reveal differences between these various types of acids which agree well in their general character with the equations derived for each. From these data he calculated  $K_R$  as about 0.35 for  $\delta$ -amino-*n*-valeric acid, 0.40 for glycylglycine, and about 0.55 for  $\omega$ -aminododecanoic acid. These values are not far from those found from solubility measurements for the same or similar substances (Chapter 11). The values from dissociation constants are of course less accurate, but the fact that such similar values of  $K_R$  are found by entirely independent methods indicates the general validity of the reasoning underlying these equations.

Recently Batchelder and Schmidt<sup>45</sup> have determined the effect of ionic strength on the dissociation of alanine, employing cells without liquid junction, of the type:  $H_2$ , solution X, AgCl, Ag. Solution X contained alanine, NaOH or HCl, and  $M^+Cl^-$ , where  $M^+$  was sodium, potassium, lithium or barium ion. From these cells, the value of  $C_H$ , the concentration of hydrogen ion in the solution, could be determined by EMF measurements<sup>46</sup>. Thus the “apparent dissociation constants” evaluated in their measurements were

$$K'_1 = \frac{C_H C_{R^+}}{C_{R^+}}, \quad K'_2 = \frac{C_H C_{R^-}}{C_{R^+}} \quad (86)$$

differing from the apparent constants (acidity constants) previously defined by the substitution of  $C_H$  for  $a_H (= C_H \gamma_H)$ . Allowing for this difference in the nature of the constants determined, their treatment follows essentially the lines of that already given. They obtained a satisfactory agreement with their accurate measurements, however, only when certain empirical parameters were introduced into the equations describing  $pK'$  as a function of ionic strength. They also studied the ionization of the trivalent amino acids aspartic acid, arginine and ornithine<sup>47</sup> as a function of ionic strength, and found characteristic differences for the different steps in ionization.

<sup>44</sup> Neuberger, A., *Proc. Roy. Soc. London (A)*, **158**, 68 (1937).

<sup>45</sup> Batchelder, A. C., and Schmidt, C. L. A., *J. Physical Chem.*, **44**, 880 (1940). See also their earlier study, *ibid.*, **43**, 1121 (1939), an extensive series of measurements on the effect of salts on  $pK'$  of glycine and alanine, using cells with liquid junction.

<sup>46</sup> See Harned and Owen, Reference 5.

<sup>47</sup> See Table 3 for their  $pK'_2$  values for these substances.

## Chapter 5

### Some Relations between Acidity and Chemical Structure

BY JOHN T. EDSALL

#### Effect of Substituent Groups on the Acidity of Neighboring Groups

In the preceding chapter we have considered the acidity of the ionizing groups in a variety of compounds. The chemist, however, frequently encounters ampholytes containing acid and basic groups of many very different types. In such cases it may be a matter of great practical importance to know whether an ampholyte with a given formula is a dipolar ion or an "uncharged molecule" in a certain solvent. Dipolar ions generally melt at very high temperatures, are more soluble in water than in most organic solvents, give solutions of high dielectric constant, and are of higher density than most organic molecules; while uncharged molecules exhibit contrasting properties in all these respects, as is shown in detail in subsequent chapters. Thus for various aspects of preparative work, in the choice of solvents for recrystallization, and in other connections, it is important to consider the relation between acidity and chemical structure.

The acidity of a group depends primarily on the nature of the group itself and secondarily on the nature of neighboring groups in the molecule. If all the neighboring groups are saturated aliphatic hydrocarbon radicals, the ionization of the acidic group is found to be nearly independent of the length of the hydrocarbon chain. This is well illustrated by the data in Chapter 4, Table 2, on the ionization constants of monoamino-monocarboxylic acids. A further example may be given from the very accurate work of Harned and his colleagues on the fatty acids and chloroacetic acid (Table 1).

The substitution of a methyl group for a hydrogen atom on the carbon atom adjoining the carboxyl group increases  $pK$  by one unit; replacing one hydrogen in this methyl group by another methyl further increases  $pK$  by 0.12 unit. Further lengthening of the hydrocarbon chain has little effect. Indeed, the  $pK$  values at  $25^\circ$  of all fatty acids except formic are

mately one hundred times as strong an acid as acetic or propionic, and all highly polar substituents produce a marked increase in acidity.

The effect of an electric charge on a substituent group has already been mentioned in Chapter 4. The ion  $^+\text{H}_3\text{N}\cdot\text{CH}_2\cdot\text{COOH}$  (pK 2.3) is a hundred times as strong an acid in water as  $\text{H}_2\text{N}\cdot\text{CH}_2\cdot\text{COOH}$  (pK 4.3): (see Chapter 4, Table 7,  $\text{pK}_\text{a}$  values), due to the charge on the proton. Similarly the  $-\text{NH}_3^+$  group in  $^+\text{H}_3\text{N}\cdot\text{CH}_2\cdot\text{COO}^-$  (pK 9.7) is a hundred times weaker than in  $^+\text{H}_3\text{N}\cdot\text{CH}_2\cdot\text{COOCH}_3$  (pK 7.7), due to the negative charge on the  $\text{COO}^-$  group. This may be denoted as the charge effect on acidity, while the effect of substituents such as  $-\text{Cl}$  may be called the dipole induction effect.

Many observed phenomena, however, cannot be explained on the basis of either of these effects. Thus the anilinium ion (pK 4.7), is a million

TABLE 1. IONIZATION CONSTANTS AND HEAT OF IONIZATION OF FATTY ACIDS AND CHLOROACETIC ACID

Substance	Ref.	$\text{pK}_{25^\circ}$	$\Delta H_{25^\circ}$ cal/mole	$\text{pK}_{\text{max}}$	$\theta(^{\circ}\text{C})$	$(\Delta C_p)_{0^\circ}$	$(\Delta C_p)_{40^\circ}$
Formic acid . . . .	(1)	3.751	-13	3.752	24.7	-34.6	-44.8
Acetic acid . . . .	(2)	4.756	-112	4.754	22.6	-32.2	-35.0
Propionic acid . .	(3)	4.874	-168	4.873	20.9	-34.7	-39.4
<i>n</i> -Butyric acid . .	(4)	4.820	-691	4.804	8.0	-32	-54
Chloroacetic acid . . . . .	(5)	2.860	-1170	2.815	-5.4	-46.8	-27.7
$\text{pK} - \text{pK}_{\text{max}} = 5 \times 10^{-5} (t - \theta)^2$ $\Delta H = -2.3 \times 10^{-4} RT^2(t - \theta)$							

(1) Harned, H. S., and Embree, J. W., *J. Am. Chem. Soc.*, 56, 1042 (1934).

(2) Harned, H. S., and Embree, J. W., *J. Am. Chem. Soc.*, 55, 652 (1933).

(3) Harned, H. S., and Embree, J. W., *J. Am. Chem. Soc.*, 55, 2379 (1933).

(4) Harned, H. S., and Sutherland, R. O., *J. Am. Chem. Soc.*, 56, 2039 (1934).

(5) Wright, D. D., *J. Am. Chem. Soc.*, 56, 314 (1934).

For values of  $\theta$  and  $\text{pK}_{\text{max}}$ , see Harned and Embree, *J. Am. Chem. Soc.*, 56, 1050 (1934). See also H. S. Harned and B. B. Owen, *Chem. Rev.*, 25, 31 (1939). H. S. Harned and R. A. Robinson, *Trans. Faraday Soc.*, 36, 973 (1940).

times as strong an acid as the methylammonium ion (pK 10.7), yet the phenyl group, like the methyl group, is a nearly non-polar substituent. The magnitude of the observed difference in acidity cannot be explained in purely electrostatic terms; a different type of effect—generally termed the resonance effect—is involved, and the explanation of this requires a detailed discussion. First, however, we shall return to a consideration of the other two effects.

### Effect of Charged Groups

Qualitatively the mechanism of the charge effect is obvious; but attempts to formulate it quantitatively have given rise to prolonged discussion. In 1923 Bjerrum<sup>1</sup> proposed a simple formulation of the effect of a charged

for many subsequent discussions. If the group, of charge  $z\epsilon$ —where  $z$  is the valence of the group, and  $\epsilon$  is the protonic charge—is regarded as immersed in a medium whose dielectric constant,  $D$ , is that of the solvent; and if it is situated at the distance  $r$  from the acidic group; then the contribution of the charged group to the electrical potential at the acidic group is

$$\psi = \frac{z\epsilon}{Dr} \quad (1)$$

The electrical work done under the influence of the potential  $\psi$ , in removing a proton from the acidic group at  $r$  to infinity is (per mole)

$$\Delta W = N\epsilon\psi = \frac{Nz\epsilon^2}{Dr} \quad (2)$$

(In all cases to be considered here,  $z = 1$ .)

These are the fundamental assumptions of Bjerrum's treatment. If now we compare the free energy of ionization of an acid containing a charged substituent group with that of a similar acid containing no such substituent, the difference may be set equal to  $\Delta W$ . Comparing, for instance, the carboxyl group in acetic acid ( $\text{CH}_3\text{COOH}$ ) and in the glycine cation ( $^+\text{H}_3\text{N}\cdot\text{CH}_2\text{COOH}$ ), we may write:

$$\Delta F^\circ(\text{acetic acid}) - \Delta F^\circ(\text{glycine}) = \Delta W = RT \ln \frac{K_{\text{GI}}}{K_{\text{HAo}}} \quad (3)$$

Combining this with equation 2 gives

$$\ln \frac{K_{\text{GI}}}{K_{\text{HAo}}} = 2.303(\text{p}K_{\text{HAo}} - \text{p}K_{\text{GI}}) = \frac{N\epsilon^2}{RTDr} \quad (4)$$

Taking  $\epsilon = 4.80 \times 10^{-10}$  esu.,  $N = 6.02 \times 10^{23}$ ,  $R = 83.5 \times 10^6$ ,  $T = 298^\circ\text{K}$ , and  $D = 78.5$ , and expressing  $r$  in Ångströms, this becomes

$$r = \frac{3.08}{\text{p}K_{\text{HAo}} - \text{p}K_{\text{GI}}} = \frac{3.08}{4.756 - 2.31} = 1.27\text{Å} \quad (5)$$

This value is obviously much too low, being less than the distance between two bonded carbon atoms. When the substituent group, however, is further removed from the acidic group, the values of  $r$  calculated from Bjerrum's formula accord much better with those calculated from known interatomic distances in the molecule<sup>2</sup>. Similar estimates of  $r$  may be obtained by comparing the dissociation of the ammonium group in an amino acid ( $\text{p}K_2$ ) with the dissociation of the same group in a corresponding amino acid ester.

It is obvious that Bjerrum's treatment can be applied to the dissociation of any group in a molecule, provided the group is sufficiently far from the acidic group to be regarded as a point charge.

The solvent is not a continuous medium; the solute molecules must form cavities of considerable size, and the dielectric constant of such a cavity is far lower than that of a medium like water. A large part of the electrical effect of the substituent must be transmitted through the molecule, that is, through a medium of low dielectric constant. Hence the effect of the substituent should be *greater* than that calculated for a medium of the dielectric constant of the solvent; thus the distances calculated from equation 4 naturally turn out to be too low. This is particularly true when substituent and ionizing group are very near; as the separation increases, more of the electrical effect is transmitted through the solvent.

These considerations formed the basis for a revision of Bjerrum's hypothesis by Kirkwood and Westheimer<sup>3</sup>, which yields far more satisfactory results than the theory in its original form. In this treatment the solute molecules are considered as cavities of dielectric constant  $D_i$ , in the solvent. The value of  $D_i$  is taken as 2, near that of the liquid paraffin hydrocarbons; but a moderate change in the value of  $D_i$  makes very little difference in the outcome of the calculations. Kirkwood and Westheimer develop this treatment for two models: (1) a spherical molecule of radius  $b$ , with an arbitrary charge distribution, immersed in a solvent of dielectric constant  $D$ ; (2) an ellipsoid of revolution, of any eccentricity—the charged substituent and the ionizing group are generally taken as located at the foci of the ellipsoid.

It is found that, for either of these models, the Bjerrum equation 2 still holds, if the dielectric constant of the solvent is replaced by an "effective dielectric constant"  $D_E$ . For the spherical model,  $D_E$  is a function of the position of the substituent and of the ionizing group within the sphere; for the ellipsoidal model, it is a function of the eccentricity of the ellipse. The specific values of  $D_E$  applicable to any particular structure have been tabulated by Kirkwood and Westheimer. In all cases, their treatment leads to reasonable values of  $r$ , entirely compatible with known data on interatomic distances and molecular configurations. Westheimer and Shookhoff<sup>4</sup> have calculated the interprotonic distances in a large number of symmetrical dicarboxylic acids from the equation, similar to 4

$$\Delta pK = \log K_1/K_2 - \log 4 = \frac{N\epsilon^2}{2.303RTD_E} \quad (6)$$

Here the term  $(-\log 4)$  must be inserted, because for acids of this type  $K_1 = 4K_2$  in the limiting case when the acidic groups are so far apart that neither influences the dissociation constant of the other<sup>5</sup>. (This corresponds to the situation formulated in equation 31 of Chapter 4 in the special



case where  $K_A = K_B = K_C = K_D$ ). Their treatment gives for oxalic acid ( $\Delta pK = 2.36$ ), for instance, 3.85 Å as the distance between the two acidic protons—a value entirely in harmony with the dimensions of the molecule as known from crystal structure determination—while the simple Bjerrum treatment gives the impossibly low value of 0.91 Å. For malonic acid ( $\Delta pK = 2.26$ ) Westheimer and Shookhoff find  $r = 4.10$  Å, where the earlier treatment gives 1.36 Å. Diethylmalonic acid gives a much larger  $\Delta pK$  value (4.48), but the calculated interprotonic distance (3.75 Å) is only slightly less than in malonic acid, since the presence of the two bulky ethyl groups increases the region of low dielectric constant adjoining the carboxyls and thereby enhances the electrostatic interactions, thus increasing  $\Delta pK$ .

A similar treatment should be applicable to the diammonium ions (hydrazinium, ethylenediammonium, etc.) whose acidic constants have been determined by Schwarzenbach<sup>6</sup> in water and alcohol-water mixtures. The aliphatic dimercaptans<sup>7</sup>, which have been studied in alcohol-water mixtures, should also furnish material for such calculations.

Westheimer and Shookhoff also calculated for the amino acids, the distances between the positive charge on the ammonium group and the negative charge on the  $-\text{COO}^-$  group; these data are discussed in detail by Kirkwood in Chapter 12.

### Dipolar Substituents

If we compare an unsubstituted acid, such as acetic acid, with a similar acid carrying a dipolar substituent, such as chloroacetic acid, the effect of the dipole of moment  $M$  on the free energy of ionization, in a homogeneous medium of dielectric constant  $D$ , should be given by the equation:

$$\Delta pK = \frac{N_e M \cos \zeta}{2.303 R T D r^2} \quad (7)$$

where  $\zeta$  is the angle between the dipole and the line joining its center to the ionizable proton, and  $r$  is the distance between the two along this line. This equation was proposed by Eucken<sup>8</sup>. It is based on the same concepts employed in Bjerrum's theory of the effect of charged substituents, and like Bjerrum's equation it gives values of  $r$  which are much too small when the substituent and the ionizable proton are close together.

Equation 7 predicts, that (for constant  $\zeta$ ) the effect of a dipolar substituent on dissociation should fall off more rapidly than that of a charged substituent. This is indeed the case, as may be seen from Table 2, in which the dissociation constants of *n*-aliphatic acids substituted with

—Br and with  $-\text{NH}_3^+$  are compared. In the  $\alpha$  position, the effect of the  $-\text{NH}_3^+$  group on the energy of dissociation is only about 30% greater than that of the C—Br dipole; in the  $\delta$  position, its effect is five times as great. Qualitatively, therefore, in this respect equation 7 is confirmed.

TABLE 2. COMPARISON OF THE DISSOCIATION CONSTANTS OF N-ALIPHATIC ACIDS SUBSTITUTED WITH BROMINE AND WITH  $\text{NH}_3$

Bromo-acids	pK	Diff. from pK of corresponding fatty acid	Amino-acids	pK	Diff. from pK of corresponding fatty acid
Bromoacetic acid.....	2.845	1.902	Glycine	2.308	2.448
$\alpha$ -Bromopropionic acid.....	2.967	1.907	Alanine	2.340	2.534
$\beta$ -Bromopropionic acid.....	4.009	0.865	$\beta$ -Alanine	3.600	1.274
$\gamma$ -Bromobutyric acid.....	4.585	0.235	$\gamma$ -Aminobutyric acid	4.230	0.590
$\delta$ -Bromovaleric acid.....	4.711	0.110	$\delta$ -Aminovaleric acid	4.270	0.551

Neuberger, A., *Proc. Roy. Soc., London A* 158, 68 (1937).

TABLE 3. INFLUENCE OF SUBSTITUENTS (R) IN DIFFERENT POSITIONS ON THE DISSOCIATION ( $\text{pK}'$ ) OF THE CARBOXYL GROUP IN  $\text{R} \cdot (\text{CH}_2)_n \text{COOH}$  AT 25°

Substituent (R)	Dipole Moment of $\text{CH}_2\text{R}$ in Debye units	Position			
		$\alpha$ ( $n=1$ )	$\beta$ ( $n=2$ )	$\gamma$ ( $n=3$ )	$\delta$ ( $n=4$ )
$\text{CH}_3$ .....	0	4.87	4.83	4.80	4.85
$\text{CH}_2=\text{CH}$ .....	0.34	4.42			
$\text{C}_6\text{H}_5$ .....	0.39	4.26			
HO.....	1.65	3.82			
HS.....	1.39	3.60			
COOR'.....	1.7	3.34	4.52		4.60
I.....	1.6	3.15	4.05	4.64	4.77
Br.....	1.8	2.86	4.01	4.58	4.72
Cl.....	1.8-1.9	2.81	4.07	4.52	4.69
COOH.....	1.7	2.92	4.24	4.36	4.42
$\text{O}_2\text{N}$ .....	3.0-3.8		3.79		
$\text{N}\equiv\text{C}$ .....	3.1-3.5	2.44			
$\text{NH}_3^+$ .....		2.31	3.60	4.03	4.21

Values for the acidity effect of the  $-\text{COOH}$  group from R. Gane and C. K. Ingold, *J. Chem. Soc.*, p. 2158 (1931); G. Schwarzenbach, *Helv. Chim. Acta*, 16, 522 (1933); for the  $-\text{NH}_3^+$  group from C. L. A. Schmidt, W. K. Appleman and P. L. Kirk, *J. Biol. Chem.*, 81, 723 (1929). Other values of  $\text{pK}$  from Landolt-Börnstein and International Critical Tables. Values of dipole moments are taken from the table in Vol. 30 (1934) of *Trans. Faraday Soc.* Two substituents which do not fit into the table should be noted: (1) the  $\text{C}=\text{O}$  group: pyruvic acid ( $\text{CH}_3\text{CO}\cdot\text{COOH}$ ) has a  $\text{pK}'$  of 2.49; the dipole moment of acetone is 2.7 D; (2) the  $\text{C}\equiv\text{C}$  triple bond: tetrolic acid ( $\text{CH}_3-\text{C}\equiv\text{C}\cdot\text{COOH}$ ) has  $\text{pK}' = 2.60$ .

It is also clear that, as equation 7 would predict, there is a close relation between the dipole moment of a substituent in a given position in an aliphatic chain and its effect on the free energy of dissociation. This was clearly pointed out by Nathan and Watson<sup>9</sup> in 1933. The effect of various dipolar substituents (and of the charged  $-\text{NH}_3^+$  groups) on the acidity

noticed in several cases between the shift in pK produced by a dipolar substituent in the  $\alpha$  position, and the dipole moment of the substituent group. This is especially true for the phenyl group, the halogens, the  $-\text{C}\equiv\text{N}$  group, and perhaps also the nitro group. For all these groups, the direction of the dipole is nearly the same—it is inclined at nearly the tetrahedral angle ( $109^\circ 28'$ ) to the bond joining the carboxyl carbon to the  $\alpha$  carbon atom. This is not true for the other groups listed in Table 3, for which the structure of the dipole is more complex, and the relation between the magnitude of the moment and the effect on pK is therefore not so clearly apparent. The effect of the  $\text{C}\equiv\text{C}$  triple bond in tetrolic

TABLE 4. INFLUENCE OF SUBSTITUENTS (R) ON THE ACIDITY (pK') OF THE AMMONIUM GROUP IN  $\text{R}(\text{CH}_2)_n\text{NH}_3^+$  AT  $20^\circ$  OR  $25^\circ$

Substituent (R)	Dipole Moment of $\text{CH}_3\text{R}$ in Debye units	Position of Substituent			
		$\alpha$ ( $n=1$ )	$\beta$ ( $n=2$ )	$\gamma$ ( $n=3$ )	$\delta$ ( $n=4$ )
$\text{CH}_3$ .....	0	10.66	10.59	10.68	10.70
$\text{CH}_2=\text{CH}$ .....	0.34	9.76			
$\text{C}_6\text{H}_5$ .....	0.39	9.38			
$\text{HO}$ .....	1.65		9.48		
$\text{H}_2\text{N}$ .....	1.23		9.98	10.62	10.86
$\text{COOR}'$ .....	1.7	7.75	9.13	9.71	10.15
$\text{COO}^-$ .....		9.72	10.19	10.40	10.69
$\text{NH}_3^+$ .....			6.98	8.58	9.32

Values for the charged  $-\text{NH}_3^+$  group (at  $20^\circ$ ) from G. Schwarzenbach, *Helv. Chim. Acta*, **16**, 522 (1933); for the  $-\text{COOR}'$  group (at  $25^\circ$ ) from J. T. Edsall and M. H. Blanchard, *J. Am. Chem. Soc.*, **55**, 2337 (1933), and from A. N. Gentile, *Proc. Roy. Soc. (London)* **A158**, 68 (1937). Other values of pK from International Critical Tables. Dipole moments as in Table 3.

For the  $^+\text{H}_2\text{N}\cdot\text{NH}_3^+$  ion, G. Schwarzenbach [*Helv. Chim. Acta*, **19**, 178 (1936)] gives  $\text{pK}_1 = -0.88$ ; for  $^+\text{H}_2\text{N}\cdot\text{NH}_2$ ,  $\text{pK}_1 = 8.10$  at  $20^\circ$ .

The pK values for the diammonium ions (substituents  $-\text{NH}_3^+$  and  $-\text{NH}_2$ ) are uncorrected for the statistical effect (see references 1 and 3). The pK of an individual  $\text{NH}_3^+$  group in  $^+\text{H}_2\text{N}\cdot(\text{CH}_2)_n\cdot\text{NH}_3^+$  is obtained by adding 0.30 to the pK values given above for the substituent  $-\text{NH}_3^+$ . Likewise the pK of an individual  $-\text{NH}_3^+$  group in  $\text{H}_2\text{N}\cdot(\text{CH}_2)_n\cdot\text{NH}_3^+$  is given by subtracting 0.30 from the values given above for the substituent  $\text{H}_2\text{N}-$ .

acid,  $\text{CH}_3\text{C}\equiv\text{C}\cdot\text{COOH}$ , is to produce a great increase in the acidity of the carboxyl ( $\text{pK}' = 2.6$ ) as compared with  $\text{CH}_3\text{CH}_2\text{CH}_2\text{COOH}$ . This appears to be out of proportion to any permanent dipole moment arising from the triple bond. Possibly the observed effect is associated with the high polarizability of the bond.

Table 4 lists similarly the effect of various substituents on the acidity of the  $-\text{NH}_3^+$  group. The effect of a substituent R on pK' in  $\text{R}\cdot(\text{CH}_2)_n\text{NH}_3^+$  is in general distinctly greater than in  $\text{R}\cdot(\text{CH}_2)_n\text{COOH}$ ; this is to be expected, since the substituent is closer to the ionizing proton in the former

stituents may be treated along the same lines adopted in their extension of the Bjerrum theory of charged substituents. Equation 7 may still be employed, again replacing  $D$  by an effective dielectric constant,  $D_E$ . The value of  $D_E$  is a function of the configuration of the molecule, but it is found to be much lower—the values of  $D_E$  (dipole) for acids in water commonly lie between 3 and 10—for dipolar than for charged substituents. This is to be expected, since the effective range of electrostatic forces arising from dipoles is much shorter than for net charges. Hence the forces from a dipole, acting on an acidic proton, are transmitted primarily through the interior of the molecule—that is, through a region of low dielectric constant. Westheimer and Shookhoff<sup>4</sup> have shown that equation 7 gives reasonable values for the dipole-proton distance for a number of acids, when the appropriate value of  $D_E$  is used; while the original treatment of Eucken, employing  $D$ , the dielectric constant of the solvent, gives values which are far too low. It has also been shown by Westheimer<sup>10</sup> that his treatment accounts quite well for the dissociation constants of many substituted benzoic acids. For other aromatic compounds, however, any purely electrostatic theory is inadequate to account for the observed effects.

Two acids of the same charge type, differing only with respect to a dipolar substituent—for instance, acetic and monochloroacetic acids—are found to have nearly the same relative acid strengths in different solvents,<sup>11</sup> even when these differ widely in dielectric constant. Such a result is in direct conflict with what would be calculated from equation 7 in its simple form; but it is entirely compatible with the treatment of Kirkwood and Westheimer, as the latter<sup>4</sup> has pointed out. The values of  $D_E$  for dipolar substituents are so low, even in water, that the transference of a molecule from water to a solvent such as ethanol produces little change in the value of  $D_E$ . Hence the effect of a dipolar substituent on ionization is nearly independent of the solvent; while that of a charged substituent, as we have pointed out in the preceding chapter, increases markedly with decrease in the dielectric constant of the solvent.

In the treatment which has been outlined here, the effect of a charged or dipolar substituent on an acid group has been expressed in terms of an electric field transmitted through space. It may also be pictured, however, as due to migration of electrons along the chain, produced by electrostatic induction arising from the substituent<sup>12</sup>. Thus in a compound such

<sup>10</sup> Westheimer, F. H., *J. Am. Chem. Soc.*, **61**, 1777 (1939).

<sup>11</sup> See for instance L. Michaelis and M. Mizutani, *Z. Physik. Chem.*, **116**, 135 (1925); M. Mizutani, *ibid.*, **118**, 318 (1925); however, acids of the same charge type with variation in the dielectric constant of the solvent, although far less than for acids of different charge types. Compare L. A. Wooten and L. R. Hummel, *J. Am. Chem. Soc.*, **57**, 2289 (1935); L. J. Minnick and M.

as  $\text{CH}_3\text{CH}_2\text{CH}_2\text{Cl}$ , the presence of the  $\text{C}^+-\text{Cl}^-$  dipole makes the carbon on the right hand side of the formula more positive than in  $\text{CH}_3\text{CH}_2\text{CH}_2\text{CH}_3$ . This carbon then attracts electrons from its neighbor on the left, which in turn becomes more positive, and the effect is thus propagated along the chain with gradually diminishing intensity. If a group containing an acidic proton is present, the charge density of the electrons on this group is diminished by the presence of such a dipole on a neighboring atom, and the proton then dissociates more readily than in an unsubstituted molecule.

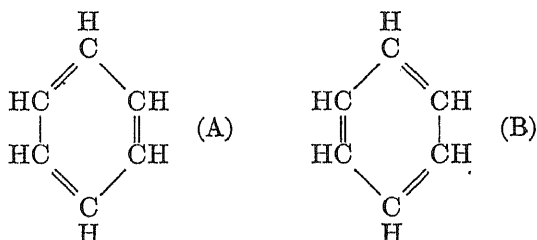
How far this inductive mechanism may be distinguished from those already discussed remains undecided; in effect, the treatment of Kirkwood and Westheimer takes account of it by regarding the molecule as a region of low dielectric constant transmitting the effects of electric charges more powerfully than the surrounding medium. An effect primarily due to shift of electrons within the molecule, however, would be practically independent of the orientation of the dipole relative to the ionizing proton; whereas equation 7, on account of the factor  $\cos \zeta$ , predicts a very large dependence on the orientation of the dipole. Too little is known, however, of the details of the structure of acids to permit further discussion of this point.

#### The Resonance Effect

Often acidity is profoundly affected by a substituent group although there is no basis for assuming a large inductive effect. A typical example is afforded by a comparison of the methyl-ammonium ion ( $\text{CH}_3\text{NH}_3^+$ , pK 10.7) and the anilinium ion ( $\text{C}_6\text{H}_5\text{NH}_3^+$ , pK 4.6). The substitution of the phenyl for the methyl group increases the acidity of the ammonium group by a factor of more than a million; yet the group, like the methyl group, is nearly non-polar.

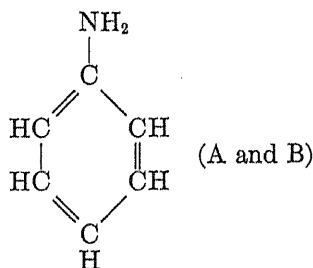
The clue to the origin of effects such as these has only been obtained in recent years, in terms of modern valence and quantum theory, largely by Pauling<sup>13</sup> and by Ingold<sup>12b</sup> and their collaborators. A full analysis of the phenomena concerned involves the principles of quantum mechanics<sup>14</sup>, but a qualitative description may be given in terms of the valence theory of G. N. Lewis<sup>12a</sup>. In Lewis' notation, most of the simplest organic compounds (aliphatic hydrocarbons, aldehydes, ketones, amines, etc.) may be described in terms of a single valence formula, in which each hydrogen atom in the molecule is surrounded by two shared electrons, and each carbon, nitrogen or oxygen atom by eight. In these cases, the "octet" principle uniquely determines the valence formula<sup>15</sup>.

In many other compounds, however, two or more formulas may be written in which the octets are all preserved. A simple and important example is benzene, in which we may write the two Kekulé formulas, A and B.



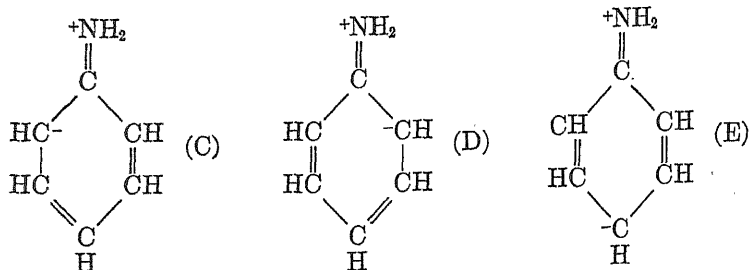
These two structures are obviously completely equivalent; one may pass into the other by a shift of bonding electrons around the ring. The actual state of the molecule does not correspond to either formula alone. Instead, it is found that all six C—C bonds in the molecule are completely equivalent, and are intermediate in character between ordinary single and double bonds. Such a structure may be shown from quantum mechanics to be necessarily more stable than either one of the component structures from which we have pictured it as being derived. In the terminology of Pauling, the molecule “resonates” between the two classical formulas; in the terminology of Ingold, the actual state is a “mesomeric state.”<sup>16</sup>

We may now offer an interpretation of the low basicity of aniline as compared with that of the aliphatic amines. The “classical” formula for the structure of aniline may be written:

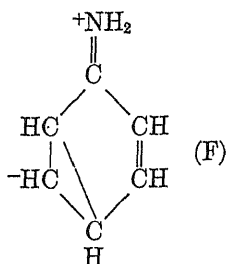


(For simplicity we have written only one of the Kekulé formulas here; the other is, of course, equally important).

Other possible formulas may be written, however, in which each atom in the aniline molecule possesses its normal octet of electrons, the unshared electron pair of the nitrogen atom being passed on to either an *ortho* or a *para* atom.



Thus the nitrogen tends to acquire a positive charge; the *ortho* and *para* carbon atoms a negative charge. A negative charge on a *meta* carbon atom is much less likely to occur, since this would involve the formation of a structure such as



which would be very unstable.

The actual state of the molecule is not given by any one of these formulas, but is determined by resonance among all of them. The C—N bond and all the C—C bonds are thus intermediate in character between single and double bonds. In this case, however, the resonating structures C, D and E are quite different from A and B, which are, of course, equivalent to each other as in the case of benzene itself. Hence there is no reason to suppose that each of the formulas written makes an equal contribution to the total configuration of the molecule. The nitrogen in aniline, as compared to that in an alkylamine, carries some excess positive charge, and the *ortho* and *para* carbons some negative charge. The magnitude of this charge depends on the relative contribution to the actual state of the molecule, made by the charged structures C, D and E, as compared with the "normal" structures A and B. The actual state of the molecule almost certainly corresponds more closely to formulas A and B than to C, D and E; since the high electric moments of the latter structures would tend to render them somewhat less stable.

If the amino group acquires a proton, however, no resonating structures

displacement involved in resonance of this type. Thus this ion is not stabilized by resonance; while the conjugate base,  $\text{C}_6\text{H}_5\text{NH}_2$  is stabilized by the resonance energy arising from the structures C, D and E. Thus the anilinium ion loses a proton far more readily than an alkylammonium ion, in which the conjugate base is not stabilized by resonance.

The difference in free energy of ionization between the anilinium ion (pK 4.6) and an alkylammonium ion (pK 10.7) may thus be attributed entirely to the resonance energy from C, D and E. Its magnitude is<sup>17</sup> (at 25°)

$$2.303 RT(10.7 - 4.6) = 8.3 \text{ kcal/mole}$$

Resonance of the type found in aniline has an important effect on the acidity of the aminobenzoic acids. In the "uncharged" form ( $\text{H}_2\text{N}\cdot\text{C}_6\text{H}_4\text{COOH}$ ) the carbon atoms *ortho* and *para* to the amino group carry an excess of negative charge due to resonance (compare forms C, D and E of aniline). For obvious electrostatic reasons, this charge tends to reduce the acidity of a carboxyl group attached to either the *ortho* or the *para* carbon. This inference is borne out by inspection of the  $\text{pK}_\text{D}$  values in Chapter 4, Table 7. These are the pK values of the carboxyl group in the uncharged form of the amino acid. While pK of benzoic acid is 4.20,  $\text{pK}_\text{D}$  for *o*-aminobenzoic acid is 4.92 and for the *para* acid 4.86; the value for the *meta* acid, however, is 4.27, almost identical with that of benzoic acid. These results are entirely in harmony with what would be expected because of resonance.

This leads to an important difference between the aminobenzoic acids with respect to the equilibrium between dipolar ions and uncharged molecules in their solutions. Consider their state when both acidic protons have been lost, that is, when they are in the form  $\text{H}_2\text{N}\cdot\text{C}_6\text{H}_4\cdot\text{COO}^-$ . When acid is added, the relative proton affinities of the two basic groups  $-\text{NH}_2$  and  $-\text{COO}^-$  determine the ratio of the two isoelectric forms. In the *o* and *p* acids, the resonance effect outlined in the preceding paragraph strengthens the basicity of the  $-\text{COO}^-$  group, thereby favoring the formation of  $\text{H}_2\text{N}\cdot\text{C}_6\text{H}_4\cdot\text{COOH}$ . In the *m* acid, this effect is absent, and there is thus a greater tendency to the formation of the dipolar ion  $^+\text{H}_3\text{N}\cdot\text{C}_6\text{H}_4\cdot\text{COO}^-$ . This is in accord with the facts; measurements of dissociation constants, of dielectric increments (Chapter 6), of electrostriction as deduced from apparent molal volume (Chapter 7) and of solubility, all indicate that the *m* acid exists in water primarily as a dipolar ion, the *o* and *p* acids primarily as uncharged molecules (Table 6).

Thus the uncharged aromatic  $-\text{NH}_2$  group decreases the acidity of an



but an aliphatic  $\text{—NH}_2$  group increases it (see  $\text{pK}_D$  of amino acid values in Chapter 4, Table 7, and  $\text{pK}_2$  values of diammonium ions in Table 4 of this chapter). Measurements of dipole moment have led to entirely concordant conclusions. Thus it was shown by Sutton<sup>18</sup> from dipole moment measurements on aniline and its derivatives that the  $\text{—NH}_2$  group is *positive* to the benzene ring, while in aliphatic amines the nitrogen is at the negative end of the C—N bond (See Ingold<sup>12b</sup>). This reversal

TABLE 5. EFFECT OF COMPLETE HYDROGENATION ON THE ACIDITY OF THE ANILINIUM AND PYRIDINIUM IONS

Substance	$\text{pK}'$	Substance	$\text{pK}'$
Aniline .....	4.62	Pyridine .....	5.21
Cyclohexylamine .....	10.61	Piperidine .....	11.13
$\Delta\text{pK}'$ .....	5.99	$\Delta\text{pK}$ .....	5.92

From N. F. Hall and M. R. Sprinkle, *J. Am. Chem. Soc.*, **54**, 3469 (1932).

TABLE 6. PHYSICAL PROPERTIES OF AMINO BENZOIC ACIDS AND ALICYCLIC AMINO ACIDS

	M.P.	pK <sub>1</sub>	pK <sub>2</sub>	K <sub>Z</sub>	Dielectric Increment	Solubility grams per liter		
						Water	Ethanol	Ether
<i>Aminobenzoic acids</i>								
<i>Ortho</i> .....	145°	2.03	4.98	0.20	>0	3.5 (14°)	107 (9.6°)	160 (7°)
<i>Meta</i> .....	174°	3.04	4.79	2.30	+41	5.9 (15°)	22 (10°)	18.1 (5.6°)
<i>Para</i> .....	187°	2.32	4.92	0.13	ca. 0	3.4 (9.6°)	113 (9.5°)	82.1 (5.8°)
<i>Aminocyclohexane carboxylic acids</i>								
2-Amino.....	273°	3.40	10.10	>10 <sup>6</sup>	24.4	Very sol.	Insol.	Insol.
3-Amino (α)...	278°	4.20	10.45	>10 <sup>6</sup>	59	Very sol.	Sl. sol.	Insol.
3-Amino (β)...	264°	4.20	10.45	>10 <sup>6</sup>	62	Very sol.	Sl. sol.	Insol.
4-Amino (α)...	260°	4.30	10.45	>10 <sup>6</sup>	62	Very sol.	Insol.	Insol.
4-Amino (β)...	235°	4.35	10.45	>10 <sup>6</sup>	61	Very sol.	Insol.	Insol.

Data on aminobenzoic acids from Beilstein's "*Handbuch*", 4th ed., Vol. XIV (See also Chapter 4, Table VII for  $\text{pK}$  and  $K_Z$  values). Data on alicyclic amino acids from J. P. Greenstein and J. Wyman, Jr., *J. Am. Chem. Soc.*, **60**, 2341 (1938). The  $K_Z$  values for these substances are roughly estimated from analogy with aliphatic amino acids of similar structure. The compounds marked  $\alpha$  were prepared from the  $\alpha$ -amino benzoic acids by hydrogenation in water; those marked  $\beta$ , by hydrogenation in HCl solution.

of the polarity of the amino group through resonance permits the interpretation of many phenomena which were previously obscure.

The possibility of resonance in aromatic rings is, of course, abolished by complete hydrogenation; correspondingly the acidity of the ammonium group in alicyclic amines is close to that of aliphatic amines. Table 5 shows the effect of hydrogenation of the ring on aniline and pyridine. Therefore the alicyclic amino acids differ radically in their physical properties from the aminobenzoic acids (Table 6). In solubility, melting point, dissociation constants, and dielectric constant increment, the former

So far as they have been investigated, the pyridine carboxylic acids, such as nicotinic acid, appear to differ in an exactly comparable way from the piperidine carboxylic acids<sup>19</sup>; and a similar contrast exists between the pyrrole carboxylic acids, and the comparable pyrrolidine compounds, such as proline.

The effect of the benzene ring on the acidity of an attached hydroxyl group is similar to its effect on the  $\text{—NH}_3^+$  group and probably due to the same mechanism. Its effect on the sulfhydryl group is similar, but considerably smaller. On a carboxyl group its effect is relatively very slight, since the atom carrying the ionizing hydrogen is not directly attached to the benzene ring, and no large resonance effect can be operative in this case (Table 7, A).

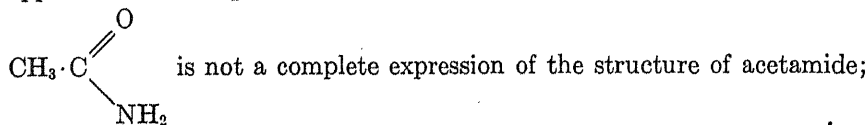
TABLE 7

A. Effect of the Phenyl Group on Acidity				
Aliphatic Acid	pK	Aromatic Acid	pK	$\Delta pK$
$\text{CH}_3\text{OH}$ .....	(17)	$\text{C}_6\text{H}_5\text{OH}$ .....	10	7.0
$\text{CH}_3\text{NH}_3^+$ .....	10.7	$\text{C}_6\text{H}_5\text{NH}_3^+$ .....	4.6	6.1
$\text{R—SH}$ .....	10.7	$\text{C}_6\text{H}_5\text{SH}$ .....	7.2	3.5
$\text{CH}_3\text{COOH}$ .....	4.75	$\text{C}_6\text{H}_5\text{COOH}$ .....	4.20	0.55
B. Effect of Replacing $\text{CH}_2$ by $\text{C}=\text{O}$ Adjoining an Atom Carrying an Acid Hydrogen				
Aliphatic Acid	pK'	Aromatic Acid	pK'	$\Delta pK'$
$\text{CH}_3\text{CH}_2\text{NH}_3^+$ .....	10.7	$\text{CH}_3\text{CO}\cdot\text{NH}_3^+$ .....	ca. -1	ca. 12
$\text{CH}_3\text{CH}_2\text{OH}$ .....	ca. 17	$\text{CH}_3\text{COOH}$ .....	4.7	12

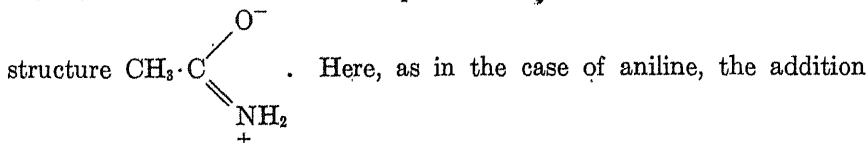
The values for the  $\text{—SH}$  compounds were estimated by G. Schwarzenbach and A. Epprecht, *Helv. Chim. Acta.*, 19, 493 (1936) by extrapolation to water of measurements in alcohol-water mixtures. Approximate values for alcohols estimated by Schwarzenbach and Epprecht from experiments of A. Unmack, *Z. Physik. Chem.*, 129, 349 (1927); 131, 371 (1928); 133, 45 (1928).

### Resonance in the Amide Group

The amides are extremely weak bases; that is, their cations are extremely strong acids. Hall and Conant<sup>20</sup> have estimated the pK value of  $\text{CH}_3\text{CONH}_3^+$  to lie between 0 and -1. Here again a resonance effect appears to be a powerful determining factor<sup>21</sup>. The classic formula



the actual state of the molecule represents resonance between this and the



<sup>19</sup> For the known properties of these compounds, see Beilstein's "Handbuch," 4th ed., Vol. XXII.

of a proton to form the ion  $\text{CH}_3\text{CO}\cdot\text{NH}_3^+$  would completely inhibit the resonance. Hence the basic form,  $\text{CH}_3\text{CONH}_2$ , is stabilized by the resonance energy, while the conjugate acid is not. The resonance energy is in this case very large<sup>22</sup>. Pauling estimates it as 21 kcal./mol. Thus an adjoining  $\text{C}=\text{O}$  group increases the acidity of the  $-\text{NH}_3^+$  group far more than does a phenyl group. Pauling estimates  $\text{pK}_\text{A}$  for  $\text{CH}_3\text{CO}\cdot\text{NH}_3^+$  in water as being of the order of  $-6$ . The experimental value of Hall and Conant differs from this by a large factor, but it is, even so, about 12 pK units more acid than an alkylammonium ion.

Comparison of the acidity of aliphatic alcohols and of carboxylic acids (Table 7 B) yields quite comparable results, which may presumably be interpreted along the same lines.

### The Strength of Certain Other Acid Groups

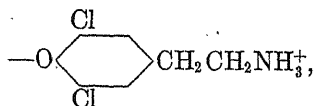
The carboxyl and ammonium groups, aliphatic and aromatic, have been discussed in some detail<sup>23</sup> in this and the two preceding chapters. We may now consider the acidity of various other groups, with particular emphasis on those found in organic ampholytes.

TABLE 8. ACIDITY OF PHENOL AND SOME CHLOROPHENOLS AT 25°

Substance	pK'	Substance	pK'
Phenol.....	9.78	2,5-Dichlorophenol.....	7.35
<i>o</i> -Chlorophenol.....	8.49	2,6-Dichlorophenol.....	6.79
<i>m</i> -Chlorophenol.....	8.86	3,4-Dichlorophenol.....	8.39
<i>p</i> -Chlorophenol.....	9.18	3,5-Dichlorophenol.....	7.93
2,3-Dichlorophenol.....	7.45	2,4,6-Dichlorophenol.....	6.42
2,4-Dichlorophenol.....	7.75		

Values from J. W. Murray and N. E. Gordon, *J. Am. Chem. Soc.*, 57, 110 (1935).

**1. The Aromatic Hydroxyl (phenolic) Group.** Phenol is a very weak acid, but the introduction of polar substituents, such as halogens, or nitro groups, into adjoining positions on the benzene ring, greatly increases its acidity (Table 8). Thus 2,6-dichlorophenol ( $\text{pK}'$  6.80) is nearly as acid as the phenolic group ( $\text{pK}'$  6.5) in the dihalogenated tyrosines (see Chapter 4, Table 7) and 2,4,6-trichlorophenol is even more acid. Thus it may be predicted that a dihalogenated tyrosine should exist in water predominantly as a dipolar ion, even after decarboxylation:



and the electric moment of such a molecule should be far greater than that of tyrosine itself<sup>24</sup>.

**2. The Sulfhydryl Group, Aliphatic and Aromatic.** Mercaptans and thiophenols are too insoluble in water to be titratable, but we owe to Schwarzenbach and his collaborators<sup>25</sup> a series of careful studies of compounds of this type in alcohol-water mixtures. Extrapolation of their results to water solution yields an estimate of 10.7 for pK in the aliphatic mercaptans, a value quite concordant with the values found for the —SH group in cysteine and cysteinyl-cysteine (Chapter 4, Table 5).

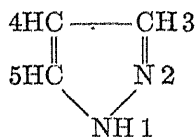
The aromatic —SH group is similarly estimated to have a pK near 7.2 in water. We may note that it follows from this value that a compound such as  $\text{H}_2\text{N} \cdot \text{CH}_2 \cdot \text{C}_6\text{H}_4 \cdot \text{SH} \rightleftharpoons {}^+\text{H}_3\text{N} \cdot \text{CH}_2 \cdot \text{C}_6\text{H}_4 \cdot \text{S}^-$  should exist predominantly in the dipolar ion form in water. Apparently no compounds of this nature have yet been prepared. Synthesis of such molecules, and study of their properties, should yield results of interest.

TABLE 9. ACIDITY OF CATIONS DERIVED FROM PYRAZOLE AND IMIDAZOLE (GLYOXALINE)

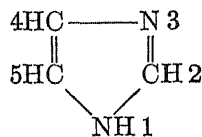
Pyrazole Derivatives	pK'	Imidazole Derivatives	pK'	pK'
Pyrazole.....	2.48	Imidazole.....	7.08	6.95†
1-Methyl pyrazole..	2.04	1-Methyl imidazole.....	7.33	
		2-Methyl imidazole.....	8.11	7.86†
3-Methyl pyrazole..	3.56	4-(or 5-) Methyl imidazole..	7.61	7.52, ‡ 7.40*
		Imidazole lactic acid.....	7.32*	
		Histamine.....	5.90*	
		2,4-Dimethylimidazole....	8.36†	
		4-(or 5-) Hydroxymethylimi-		
		dazole.....	6.38†	
		2-Pyridylimidazole.....	6.39†	
		4-Pyridylimidazole.....	6.00†	

The values marked \* were obtained by M. Levy (*J. Biol. Chem.*, 109, 361 (1935)) from potentiometric measurements at 20°. Values marked † are from the potentiometric measurements at 25° of A. H. M. Kirby and A. Neuberger, *Biochem. J.*, 32, 1146 (1938); their pK values are extrapolated to zero ionic strength. The other values were obtained by potentiometric measurements by Deussen, *Ber.*, 39, 1831 (1906); Backer, *Amer. J.*, 273, 214 (1933); or by measurements by Deussen, *Ber.*, 39, 1831 (1906). Levy gives for the —COOH group in imidazole lactic acid pK' = 6.38; in histamine pK' = 9.70 at 30°.

**3. The Imidazole and Pyrazole Groups.** These two isomeric ring structures are interesting in that they and their derivatives differ so widely in acidity (Table 9).



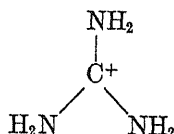
Pyrazole



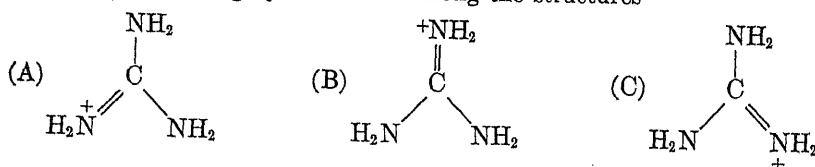
Imidazole (glyoxaline)

The difference in acidity between the imidazolium and the pyrazolium ions is very striking, and no clear explanation of it appears to be available at present. The great influence of methyl groups in certain positions on the acidity of both these ring structures is also apparent. Undoubtedly there is a large amount of resonance in both imidazole and pyrazole, so that the structures as written are not to be regarded as rigidly fixed, but there is a "blending" of the single and double bonds into bonds of intermediate character.

4. The Guanidonium (Guanidinium) Group. The base guanidine is extremely strong; it acquires a proton readily at any pH below 14 or thereabouts, and is converted into the guanidinium ion.



This ion, as shown here, is completely symmetrical with respect to the three  $\text{—NH}_2$  groups; all the three C—N distances are equal, and intermediate in character between double and single bonds; the angle between any two adjoining C—N bonds is  $120^\circ$ . This structure has been supported by quantum mechanical reasoning<sup>26</sup> and demonstrated experimentally from studies of x-ray diffraction data<sup>27</sup> and of Raman spectra<sup>28</sup>. It may be thought of as arising by resonance among the structures



Owing to the complete equivalence of the three forms, the resonance energy is high, and the ion is very stable (compare the discussion of benzene earlier in this chapter). Removal of a proton diminishes the symmetry and stability; thus the guanidinium ion in an extremely weak acid ( $\text{pK ca. 14}$ ); in other words, guanidine is a base about as strong as sodium hydroxide.

Davis and Elderfield<sup>29</sup> determined the acidity of a large number of substituted guanidinium ions and found all the monoalkyl derivatives to be as weak acids as the guanidinium ion itself. The same was true of the symmetrically trisubstituted guanidines, and also of the disubstituted derivatives in which both alkyl groups were on the same nitrogen atom. The

phenyl guanidines, however, and the symmetrical dialkyl derivatives, were found to have  $pK'$  values between 10 and 11 (Table 10). An interpretation of this remarkable behavior in terms of resonance effects has been offered by Pauling<sup>30</sup>.

Arginine, being a monoalkyl-substituted guanidine, might be expected to have a  $pK$  value near 14 for this group. The reported values of 12.5–13.2 indicate an influence of the  $\alpha$  amino and the carboxyl group which tends to increase the acidity of the guanidinium group in arginine.  $pK'$  of creatinine is 4.78 at 25<sup>°</sup><sup>31</sup>, showing the profound increase in the acidity of the guanidinium ion produced by ring formation involving a highly polar (C=O) group.

The carboxyl group in creatine ( $pK'$  2.62 at 30<sup>°</sup>, according to Cannan and Shore<sup>31</sup>) is more than one hundred times as strong an acid as in propionic acid. This represents the influence of a methylated guanidinium group with a positive charge, attached to the  $\alpha$ -carbon atom.

TABLE 10. ACIDITY OF GUANIDINIUM IONS AT 25<sup>°</sup>

Substance	Ref.	$pK'$	Substance	Ref.	$pK'$
Guanidine.....	(2)	13.65	N,N'-Dimethylguanidine..	(1)	10.29
Methylguanidine.....	(1)	ca. 14	N,N'-Diethylguanidine...	(1)	10.30
n-Butylguanidine.....	(1)	ca. 14	N,N'-di-n-Amyl guani-		
			dine.....	(1)	10.25
n-Heptylguanidine.....	(1)	ca. 14	N,N', N''-Trimethylguani-		
			dine.....	(1)	ca. 14
Benzyl guanidine.....	(1)	ca. 14	Phenylguanidine.....	(1)	10.77
N,N-Dimethylguanidine..	(1)	ca. 14	N,N'-Diphenylguanidine..	(2)	10.12
Creatine.....	(3)	ca. 14	Triphenylguanidine.....	(2)	9.10

(1) Davis, T. L., and Elderfield, R. C., *J. Am. Chem. Soc.*, **54**, 1499 (1932).

(2) Hall, N. F., and Sprinkle, M. R., *J. Am. Chem. Soc.*, **54**, 3470 (1932).

(3) Cannan, R. K., and Shore, A., *Biochem. J.*, **22**, 920 (1928).

**5. The Phosphoric Ester Group.** Phosphoric acid in ester linkage plays an important part in many naturally occurring ampholytes, such as phosphocreatine, lecithin, and also coenzymes I and II (the phosphopyridine nucleotides). The effect of esterification is in general to increase the acidity of both of the remaining free acid groups, as shown in Table 11 for the hexosephosphates.

The electrometric titration of phosphocreatine<sup>32</sup> shows the presence of a  $pK'$  value at about 2.7 (presumably the carboxyl group) and another at 4.6 (presumably the second acid group of the phosphoric acid radical). There is also indication of a  $pK'$  value less than 2 (first acid group of phosphoric radical) and of one near 12 (the guanidinium group).

The strength of the acidic hydrogen in the phosphoric ester group of lecithin and cephalin is not exactly known, but it is almost certainly very

strong ( $pK < 2$ ). Lecithin is certainly a dipolar ion with the positive charge on the nitrogenous radical and the negative charge on the phosphate group. The structure of cephalin is still open to question<sup>33</sup>

TABLE 11.  $pK'$  VALUES OF PHOSPHORIC ACID AND HEXOSEPHOSPHORIC ACIDS (from A. Harden, "Alcoholic Fermentation," Fourth Ed., New York and London, 1932)

	$pK_1'$	$pK_2'$
Phosphoric Acid.....	1.99	6.81
Hexosemonophosphate:		
Neuberg ester.....	0.97	6.11
Embden ester.....		6.12
Robison ester.....	0.94	6.11
Hexosediphosphate:		
Harden and Young ester.....	1.48	6.29
Lohmann ester.....		6.3

Data from Meyerhof and his collaborators.  $pK_1$  of  $H_2PO_4$  has been very accurately determined by L. F. Nims [*J. Am. Chem. Soc.*, 56, 1110 (1934)] as 2.124 at 25°; and  $pK_2$  as 7.205 at 25° [L. F. Nims, *J. Am. Chem. Soc.*, 55, 1946 (1933)].

6. Quaternary Ammonium Groups in Dipolar Ions. All ammonium groups of this type are of course exceedingly weak acids, with  $pK$  generally of the order of magnitude of 14 or more. Hence any molecule containing a group of this type, together with an acidic group of the uncharged type (such as  $-COOH$  or  $-OH$ ) should be a dipolar ion. Thus while the aminophenols exist almost entirely in the form of uncharged molecules, the *N*-trimethyl aminophenols are dipolar ions of high electric moment.  $pK'$  for *N*-trimethyl *o*-amino-phenol is 7.42 for the  $-OH$  group, for the *meta* and *para* compounds about 8<sup>34</sup>. The dipolar ionic nature of these compounds has been verified from dielectric constant measurements by Devoto (see Chapter 6, references given in Table 3).

#### The Effect of the Peptide Linkage on Acidic Dissociation

The influence of this very important linkage, uncomplicated by the effects of other substituents, is best seen in acetyl-glycine ( $pK'$  3.60), which is more acid by 1.2 pH units than the corresponding fatty acid, *n*-valeric acid. Formyl glycine ( $pK'$  3.42) is definitely more acid than acetyl glycine although a substitution of H for a  $-CH_3$  group in a position as far as this from an ionizing carboxyl group generally has a negligible effect. This shows the great sensitivity of the resonating  $-CONH$  group to change in the nature of the directly attached atom.

The effect of the peptide linkage, and of some other substituents, on the dissociation of an adjoining carboxyl group, is shown in Table 12, taken from the work of Zief and Edsall<sup>35</sup>.

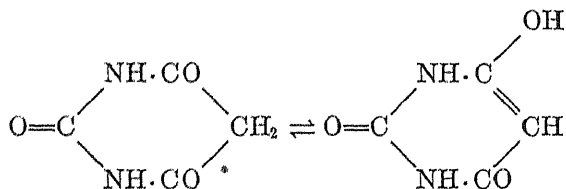
TABLE 12. EFFECT OF THE PEPTIDE LINKAGE AND OF OTHER SUBSTITUENTS ON THE DISSOCIATION OF THE CARBOXYL GROUP

Name	Formula	pK' (COOH)
<i>n</i> -Valeric acid	CH <sub>3</sub> CH <sub>2</sub> CH <sub>2</sub> CH <sub>2</sub> COOH	4.80
$\delta$ -Chlorovaleric acid	ClCH <sub>2</sub> CH <sub>2</sub> CH <sub>2</sub> CH <sub>2</sub> COOH	4.69
Levulinic acid	H <sub>3</sub> CCOCH <sub>2</sub> CH <sub>2</sub> COOH	4.59
Glycine (uncharged)	H <sub>2</sub> NCH <sub>2</sub> COOH	(4.30)
$\delta$ -Aminovaleric acid	<sup>+</sup> H <sub>3</sub> NCH <sub>2</sub> CH <sub>2</sub> CH <sub>2</sub> CH <sub>2</sub> COOH	4.21
Hydantoic acid	H <sub>2</sub> NCONHCH <sub>2</sub> COOH	3.80
N-Carbethoxyglycine	C <sub>2</sub> H <sub>5</sub> COONHCH <sub>2</sub> COOH	3.65
Malonic monoamide	H <sub>2</sub> NCOCH <sub>2</sub> COOH	3.64
Acetylglycine	CH <sub>3</sub> CONHCH <sub>2</sub> COOH	3.60
Acetoacetic acid	H <sub>3</sub> CCOCH <sub>2</sub> COOH	3.58
Glycylglycylhydantoic acid	H <sub>2</sub> NCONHCH <sub>2</sub> CONHCH <sub>2</sub> COOH	3.54
Glycylglycine (uncharged)	H <sub>2</sub> NCH <sub>2</sub> CONHCH <sub>2</sub> COOH	(3.46)
Formylglycine	HCONHCH <sub>2</sub> COOH	3.42
Chloroacetylglycine	ClCH <sub>2</sub> CONHCH <sub>2</sub> COOH	3.37
Glycylglycine (charged)	H <sub>2</sub> NCH <sub>2</sub> CONHCH <sub>2</sub> COOH	3.14
Pyruvic acid	CH <sub>3</sub> C(=O)COOH	2.49
Glycine (charged)	H <sub>2</sub> NCH <sub>2</sub> COOH	2.31

From M. Zief and J. T. Edsall, *J. Am. Chem. Soc.*, 59, 2247, (1937), Table III.

### The Acidity of Some Ring Structures

An interesting and important type of ring structure is that of hydantoin,  $\text{NH}\cdot\text{CO}\cdot\text{NH}\cdot\text{CO}\cdot\text{CH}_2$ . This molecule is an acid of  $\text{pK}'$  9.12 at 25°, as determined by Zief and Edsall<sup>35</sup>. Studies of derivatives in which some of the hydrogens have been replaced by alkyl or aryl substituents show that the acidic hydrogen is situated on the nitrogen atom between the two C=O groups<sup>36</sup>. Similarly the =NH group in succinimide is an acid ( $\text{pK}' = 9.56$ )<sup>37</sup>. In barbituric acid ( $\text{pK}' = 3.95$ ), Wood<sup>38</sup>



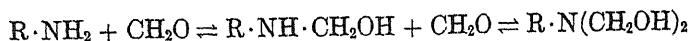
has shown that the acidity arises principally from the hydrogen of the CH<sub>2</sub> group (after enolization).

The ionization of ascorbic acid ( $\text{pK}'_1 = 4.12$ ,  $\text{pK}'_2 = 11.51$ ) has been studied by Kumler and Daniels<sup>39</sup> and that of tetronic acid and some of its derivatives by Kumler<sup>40</sup> who has interpreted the data in terms of resonance.

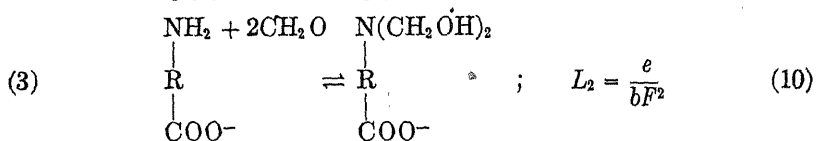
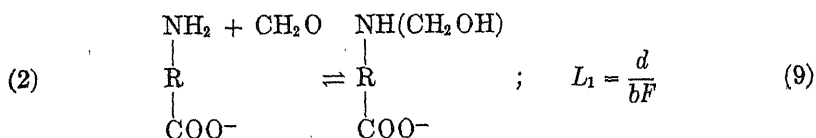
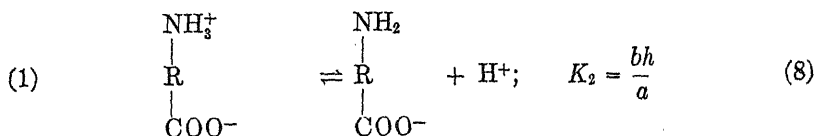


### The Effect of Formaldehyde upon the Titration Constants of Amino Acids

It has long been known that formaldehyde exerts a marked effect upon certain portions of the titration curves of amino acids. Schiff<sup>41</sup> showed that some amino acids when titrated in formaldehyde neutralized an equivalent of alkali, using phenolphthalein as an indicator. Sørensen<sup>42</sup> developed the method and applied it widely for the quantitative study of protein hydrolysis by enzymes, and to other systems. The very extensive practical use which has since been made of the method cannot be discussed here, but we shall present a brief discussion of some of the underlying factors involved. The situation is complex; we shall follow the treatment of Levy<sup>43</sup>, which appears to provide at present the most satisfactory method for describing the facts. He assumes that an un-ionized amino group may combine with either one or two molecules of formaldehyde, but that a charged ammonium group does not combine with it at all.



A monoamino-monocarboxylic acid thus combines with formaldehyde only on the alkaline side of its isoelectric point. Three equilibria must be considered.



$K_2$  is the second dissociation constant of the amino acid;  $L_1$  and  $L_2$  are association constants between the amino acid anion and one and two moles of formaldehyde respectively.

If  $\alpha$  be defined as the fraction of the total amino acid present as one form

<sup>41</sup> Schiff H. *Annalen*. 319. 59 (1901): 325, 348 (1902).

or another of the anions, and complete dissociation of salts is assumed, we obtain

$$\frac{\alpha}{1 - \alpha} = \frac{b + d + e}{a} \quad (11)$$

From equations 8, 9, 10, and 11 we obtain

$$\text{pH} = \text{pK}_2 - \log (1 + L_1 F + L_2 F^2) + \log \frac{\alpha}{1 - \alpha} \quad (12)$$

TABLE 13. CONSTANTS DESCRIBING THE FORMOL TITRATION OF AMINO AND IMINO ACIDS AT 30°

Amino Acid	pK	L <sub>1</sub> *	L <sub>2</sub> *	pKL <sub>1</sub> †	9% CH <sub>2</sub> O pG <sub>f</sub> †	pG <sub>f</sub> †
Glycine.....	9.60	60	290	6.65	5.70	5.92
dl-Alanine.....	9.72	22	75	7.81	6.86	6.96
l-Leucine.....	9.50	16	35	7.87	6.92	
dl-Valine.....	9.50	22	12	8.20	7.25	7.47
dl-α-Aminophenylacetic acid.....	8.84	13	77	6.90	5.95	
l-Phenylalanine.....	8.99	16	23	7.57	6.62	6.80
l-Tyrosine.....	9.07	10	5	8.45	7.50	
				pKL <sub>1</sub>		
l-Tryptophane.....	9.27	83		7.35	6.88	
l-Proline.....	10.30	112		8.25	7.78	
l-Hydroxyproline.....	9.56	79		7.66	7.19	
dl-Sarcosine.....	10.06	151		7.88	7.41	
				pKL <sub>2</sub>		
d†-Lysine,§ pK <sub>2</sub> .....	10.56	240	309	8.10	7.15	
d†-Arginine,§ pK <sub>2</sub> .....	8.91		6 × 10 <sup>5</sup>			
l-Histidine, pK <sub>3</sub> .....	9.17				7.90	
d†-Glutamic acid, pK <sub>1</sub> .....	9.32	22	24	7.87	6.91	
l-Aspartic acid, pK <sub>3</sub> .....	9.83	26	37	8.16	7.21	

\* L<sub>1</sub> is the association constant of the amino or imino acid anion for 1 molecule of formaldehyde. L<sub>2</sub> is the association constant of the amino acid anion for 2 molecules of formaldehyde.

† This constant will differ from pK<sub>n</sub> — log L<sub>2</sub>, because it is derived for an approximation formula and is read directly from the curves.

‡ Data of M. S. Dunn and A. Loshakoff, *J. Biol. Chem.*, **113**, 359 (1936).

§ For a complete description of the behavior of these compounds in formaldehyde, see M. Levy, *J. Biol. Chem.*, **99**, 767 (1932-33).

(Data from M. Levy and D. E. Silberman, *J. Biol. Chem.*, **118**, 723 (1937)).

The application of this equation to the formol titration requires, either that  $F$  be replaced by the exact relationship in which  $C$  is the formaldehyde added to the mixture,

$$F = C - d - 2e \quad (13)$$

or that an approximation be introduced. The use of the exact relationship leads to a complicated relationship and an approximation is readily available. If the total amino acid concentration is small compared to the

formaldehyde combined and to substitute for  $F$  in equation 13 the total added formaldehyde.

Therefore, at a constant and sufficiently high formaldehyde concentration the second term on the right of equation 12 becomes constant and  $\text{pK}_2 - \log(1 + L_1F + L_2F^2)$  is equal to the pH at the mid-point of titration of the amino acid. This titration constant may be designated  $\text{pG}_t$  and defined by equation 14.

$$\text{pG}_t = -\log(K_2 + K_2L_1F + K_2L_2F^2) \quad (14)$$

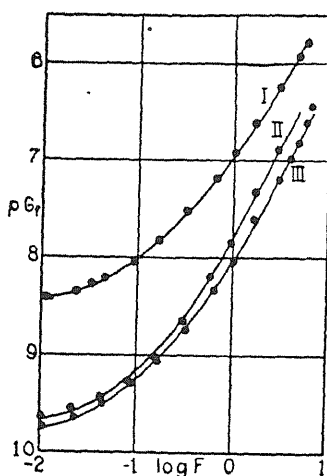


Figure 1

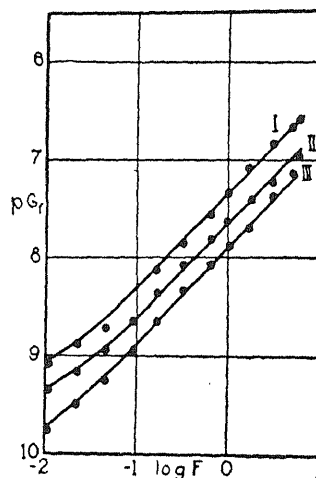


Figure 2

Figure 1. The ordinate scale represents the apparent dissociation constants of the amino acids; the abscissa, the logarithms of the formaldehyde concentrations in moles per liter. Curve I, valine (I has been subtracted from the  $\text{pG}$  values for plotting); Curve II, *dl*-alanine; Curve III, aspartic acid.

Figure 2. Ordinate and abscissa as in Figure 1. Curve I, tryptophane; Curve II, sarcosine; Curve III, hydroxyproline.

From Levy, M., and Silbermann, D. E., *J. Biol. Chem.*, 118, 724 (1937).

The titration constants obtained from hydrogen electrode measurements on amino acids in formaldehyde solutions can be described in terms of these equations. It is remarkable, however, that the association constants  $L_1$  and  $L_2$  differ widely from one amino acid to another, even when the amino acids differ only with respect to the size of the hydrocarbon radical in the side chain. The reason for this extreme degree of individuality is not yet apparent, but the fact is unmistakable. The results obtained are summarized in Table 13, taken from the work of Levy and

sarcosine,  $L_2$  is zero; that is, these acids (with only one N—H linkage in the un-ionized imino group) combine with only one molecule of formaldehyde. The same is also true of tryptophane. The difference between substances of this type and the primary amino acids is shown in Figs. 1 and 2, in which the observed logarithmic titration constant is plotted against the logarithm of the formaldehyde concentration.

In all cases the effect of formaldehyde at constant pH is to shift the amino acid equilibrium in the direction of the anionic forms. At constant pH, from equation (12), we have

$$\log \frac{\alpha}{1 - \alpha} = \log (1 + L_1 F + L_2 F^2) + \text{a constant} \quad (15)$$

where  $\alpha$  is the total fraction of the amino acid in the anionic form. On the other hand, the pH at which an indicator such as phenolphthalein changes color is not appreciably affected by formaldehyde. These facts are the basis of the formaldehyde titration. For accurate titration, however, it is necessary to take account of the individual differences between the different amino acids, and adjust the conditions accordingly.

Dunn and Loshakoff<sup>44</sup> and Dunn and Weiner<sup>45</sup> have described a formaldehyde titration method, employing the glass electrode, which appears to give results of very high accuracy for amino acid solutions.

<sup>44</sup> Dunn, M. S., and Loshakoff, A., *J. Biol. Chem.*, **113**, 359, 691 (1936).

<sup>45</sup> Dunn, M. S., and Weiner, J. G., *J. Biol. Chem.*, **117**, 381 (1937).

## Chapter 6

### Dielectric Constants and Dipole Moments of Dipolar Ions

BY JOHN T. EDSALL

#### Definition and Measurement of the Dielectric Constant

The distinctive properties of dipolar ions are most clearly revealed in the high dielectric constants of their solutions, which are the direct consequence of their extraordinarily large electric moments. Amino acids, peptides, proteins and phospholipoids give solutions of higher dielectric constant than can be found in any other known substances. Among the amino acids and peptides, the relations between dielectric constant and chemical structure are extraordinarily simple and regular.

The dielectric constant,  $D$ , may be defined in three equivalent ways.

1. The Law of Force Between Charged Bodies. The force,  $F$ , between two point charges,  $q_1$  and  $q_2$ , separated by a distance,  $r$ , is inversely proportional to the dielectric constant of the medium surrounding the charges

$$F = \frac{q_1 q_2}{Dr^2} \quad (1)$$

$D$  is defined as unity for a vacuum, and is greater than unity for all other media.

2. If  $C_0$  is the capacity of a condenser *in vacuo*<sup>1</sup>, and  $C$  its capacity in another medium, then the dielectric constant of the medium,

$$D = C/C_0 \quad (2)$$

This relation furnishes the basis for most experimental measurements of the dielectric constant.

3. It was shown by Maxwell that the velocity of propagation of an electromagnetic wave in a medium of dielectric constant,  $D$ , is inversely proportional to  $\sqrt{D}$ . Since this velocity is also inversely proportional to the refractive index,  $n$ , of the medium, we obtain Maxwell's relation

$$D = n^2 \quad (3)$$

light of much higher frequency. Thus equation 3 is generally valid only when  $D$  and  $n$  are measured *at the same frequency*. This emphasizes the important fact that the dielectric constant is a function of the frequency of the electric field. At sufficiently low frequencies (long wave lengths), however,  $D$  becomes independent of the frequency, and identical with the value determined in a static field. It is this low frequency dielectric constant with which the subject matter of this chapter is concerned<sup>2</sup>. In the study of proteins, however, the variation of dielectric constant with frequency is of major importance, and will require detailed discussion (Chapter 22).

Dielectric constants of various liquids at 20° are given in Table 1. Non-polar liquids like benzene and hexane have very low values, near 2. Water and hydrocyanic acid have the highest values known for pure

TABLE 1. DIELECTRIC CONSTANTS AND DIPOLE MOMENTS OF CERTAIN IMPORTANT SUBSTANCES

Substance	Dielectric constant $D$ at 20°	Dipole moment $\times 10^{18}$ e.s.u.
Vacuum.....	1.00	—
Hexane.....	1.87	0
Octane.....	1.96	0
Benzene.....	2.28	0
Toluene.....	2.39	0
Diethyl ether.....	4.33	1.15
Chloroform.....	5.05	1.15
Acetone.....	21.4	2.75
Ethanol.....	24	1.7
Methanol.....	33	1.7
Water.....	80	1.9
Hydrocyanic acid.....	116	2.6
2.5 $M$ Glycine in water.....	137	(15)

(See Hildebrand, J. H., "Solubility," 2nd ed., New York, Reinhold Publishing Corp., 1936, Chapter IV.)  
The value for glycine solution is estimated from data concerning this substance, the references for which may be found in Table 3.

<sup>2</sup> Taken from Edsall, J. T., in Schmidt, C. L. A., *The Chemistry of the Amino Acids and Proteins*, C. C. Thomas, Springfield, Ill., 1938, p. 878.

liquids. An aqueous solution saturated with glycine, however, has a much higher value still, and other amino acids and peptides give solutions of dielectric constant far higher than this.

#### Relation of Dielectric Constant to Molecular Structure: Polar and Non-Polar Molecules

The nature of the relation between dielectric constant and molecular structure has been elucidated especially by the work of Debye<sup>3</sup> (See also<sup>4</sup>). The fundamental assumption involved has been made from the time of

<sup>3</sup> The dielectric constants of amino acid and peptide solutions are independent of frequency for all frequencies up to  $10^8$  cycles per second, or even above this. This is true of most non-colloidal liquids, except for a few alcohols and certain other molecules. Recent work has shown, however, that the dielectric constants of amino acid and peptide solutions decrease with increasing frequency if sufficiently high frequencies are

Faraday: that the molecules of a dielectric medium become electrically polarized by the applied field. Thus, in the medium between the plates of a charged condenser, each element of volume acquires an electric moment; the molecules act as dipoles, the negative end of the dipoles pointing toward the positive plate of the condenser. It is the field arising from these dipoles which increases the capacity of the condenser.

The electric moment induced in the medium by the external field may arise in either of two ways.

**1. The Distortion Effect.** Since all molecules are made up of positively charged nuclei and negatively charged electrons, the electrons are attracted in the direction of the positive field, the nuclei in the opposite direction. (The actual displacement of position is far greater for the electrons than for the nuclei, because of their much smaller mass.) Through the distortion thus produced, the molecules become electric dipoles, the moment  $m$  of the dipoles being proportional to the intensity,  $F$ , of the field acting upon the molecule<sup>5</sup>  $m = \alpha_0 F$ . The constant of proportionality,  $\alpha_0$ , depends on the ease of displacement of the electrons in the molecule, and is known as the *polarizability*. All molecules, of course, are polarizable, and, therefore, give rise to electric moments by this mechanism in an electric field. These induced moments are practically independent of temperature.

**2. The Orientation Effect.** Many molecules however, are dipoles even in the absence of an external field. A simple and familiar example is water which has the form of a symmetrical triangle, the angle between the two O—H bonds being very nearly  $105^\circ$ . The oxygen nucleus (charge  $+8e$ )<sup>6</sup> tends to draw around it not only its own electrons but also those of the hydrogens, thus becoming the center of a region of net negative charge; the hydrogen nuclei being left with a residual positive charge. Thus the whole structure is a dipole, the negative end of which is near the oxygen nucleus, the positive end on a line half way between the protons. We designate the moment of this permanent dipole by the symbol  $\mu$ .

In the absence of an external field, these dipoles are distributed at random, owing to their thermal agitation, and the net moment of the system per unit volume is zero. In the presence of an electric field, however, the permanent dipoles become oriented, the degree of orientation produced by a given field strength being greater at low temperatures where thermal agitation is small. Debye has shown, for gases, in which the molecules are far apart, that the average moment per molecule due to this

<sup>5</sup> This internal field intensity,  $F$ , is not in general identical with the intensity  $E$ , which would be determined by an instrument (such as an electrometer) for determining the rate of fall of potential across the field.

orientation effect, in the direction of the field  $F$ , is equal to  $\mu^2 F/3kT$ ; that is, it is proportional to the square of the permanent dipole moment of the molecule, and to the reciprocal of the absolute temperature<sup>7</sup>.

The total average moment per molecule,  $\bar{m}$ , in the direction of the field is the sum of the distortion and orientation effects:

$$\bar{m} = \left( \alpha_0 + \frac{\mu^2}{3kT} \right) F \quad (4)$$

and the moment per mole of gas is  $N$  times as great.

If the field is of unit strength (1 esu = 300 volts/cm), then the total moment per mole in the direction of the field, multiplied by the factor  $4\pi/3$ , is known as the polarization,  $P$ . For gases, the polarization is very simply related to the dielectric constant,  $D$ , by the formula:

$$P = \frac{4\pi}{3} N \left( \alpha_0 + \frac{\mu^2}{3kT} \right) = \frac{D - 1}{D + 2} V \quad (5)$$

when  $V$  is the molal volume of the gas. This equation, for substances in which  $\mu = 0$ , was first derived by Clausius and Mosotti; the form given here, which applies also to polar gases, was given by Debye. It will be seen from equation 5 that the dielectric constant or the polarization of a non-polar gas, like  $H_2$  or  $N_2$ , at constant volume, is independent of temperature. In substances such as dipolar ions, on the other hand, the moment induced by distortion,  $\alpha_0 F$ , is very small compared to the permanent moment. In such cases, the polarization, at constant temperature, is directly proportional to the *square* of the permanent moment.

If a small number of polar molecules be dissolved in a large amount of a non-polar solvent, the polar molecules are generally almost as independent of one another as if they were in a gas. For such systems, Debye<sup>3</sup> has developed an extended form of equation 5, which permits the calculation of the electric moment of a polar solute from measurements of the dielectric constant of its solution in a non-polar solvent<sup>8</sup>.

The order of magnitude of the electric moments to be expected in simple molecules may be readily estimated. The charge on the electron ( $-\epsilon$ ) is  $4.80 \times 10^{-10}$  esu., and the distances between atoms in a molecule are of the order of magnitude of one Ångstrom unit ( $= 1 \times 10^{-8}$  cm). A dipole containing charges  $+\epsilon$  and  $-\epsilon$ , separated by 1 Å, should, therefore, have a moment  $(4.80 \times 10^{-10}) \times 10^{-8} = 4.80 \times 10^{-18}$  electrostatic units. Experimentally determined values for a number of molecules are given in Tables 2 and 3. The values range from 0 (for the hydrocarbons) up to  $8.6 \times 10^{-18}$  esu. for urea in dioxane. The latter is as high as any value that has been



TABLE 2. DIELECTRIC INCREMENTS AND DIPOLE MOMENTS OF CERTAIN MOLECULES  
(All values are at 25°)

Substance		$\delta = \frac{dD}{dC}$ in water	Dipole moment, $\mu$ in Debye units, $D$ (1 $D = 10^{-18}$ e.s.u.)
Methanol.....	(a)	-1.4	1.65
Ethanol.....	(a)	-2.6	1.7
<i>n</i> -Propanol.....	(a)	-4.0	1.7
<i>iso</i> -Propanol.....	(a)	-4.3	1.7
Butanol (tertiary).....	(a)	-6.3	1.7
Acetone.....	(a)	-3.2	2.75
Ether.....	(b)	-7.1	1.1
Dioxane.....	(c)	-8.3	0.0
Glycol.....	(a)	-1.8	2.3
Glycerol.....	(a)	-2.6	?
Mannitol.....	(a)	-2.6	?
Sucrose.....	(a)	-7.5	?
Aniline.....	(4)	-7.6	1.52
Phenol.....	(4)	-6.6	1.65
Methyl acetate.....	(4)	-5	1.75
Pyridine.....	(4)	-4.2	2.2
Nitromethane.....	(4)	-2.0	3.05
Acetonitrile.....	(4)	-1.7	3.3
Acetamide.....	(5)	-0.8	3.8
Dimethylurea (Symmetrical).....	(4)	+3.0	4.8
Urea.....	(6) (d)	+2.7	8.6? (9)
Thiourea.....	(7) (d)	+4	7.6? (9)
Malonamide.....	(5)	+4.3	?
Glycine anhydride.....	(5)	-10	?
Glycine.....	(8)	+22.6	about 15
$\alpha$ -Amino- <i>n</i> -valeric acid.....	(8)	+22.6	about 15
Diglycine.....	(8)	+71	about 25

Values of  $\mu$  taken from table in *Trans. Faraday Soc.*, 30, 1934, except those for amino acids, which are estimated by methods discussed in text.

(a) Calculated  $\delta$ -values are based on the dielectric constant data of Akerlöf (1).

(b) Based on data given by Falkenhagen (2, p. 149).

(c) Data of Wyman, given by Scatchard and Benedict (3).

(d) Kumler, W. D., and Fohlen, G. M. *J. Am. Chem. Soc.*, 64, 1944 (1942) obtain from very careful measurements the moments 4.56 for urea, and 4.89 for thiourea.

Taken from Edsall, J. T., in Schmidt, C. L. A., "The Chemistry of the Amino Acids and Proteins," C. C. Thomas, Springfield, Ill., 1938, p. 882.

(1) Akerlöf, G., *J. Am. Chem. Soc.*, 54, 4125 (1932).

(2) Falkenhagen, H., *Elektrolyte*, Leipzig (1932).

(3) Scatchard, G., and Benedict, M. A., *J. Am. Chem. Soc.*, 58, 837 (1936).

(4) Devoto, G., *Ber. chem. Ges.*, 67, 1925 (1934).

(5) Devoto, G., *Gazz. Chim. Ital.*, 61, 847 (1931).

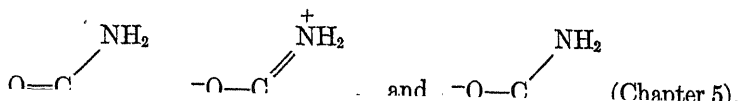
(6) Wyman, J., Jr., *J. Am. Chem. Soc.*, 55, 4116 (1933).

(7) Devoto, G., *Gazz. Chim. Ital.*, 60, 520 (1930).

(8) Wyman, J., Jr., and McMeekin, T. L., *J. Am. Chem. Soc.*, 55, 908 (1933).

(9) Bergmann, E., and Weizmann, A., *Trans. Faraday Soc.*, 34, 783 (1938).

reported for any substance soluble in a non-polar solvent, and reflects the electrical asymmetry produced by resonance between the structures



### Dipolar Ions

It is impossible to determine the electric moments of substances like the amino acids by the methods mentioned above. They decompose on melting and are insoluble in all non-polar solvents. Information of the utmost importance has been obtained, however, by dissolving them in polar solvents (such as water or one of the alcohols) and studying the dielectric constant of the solution<sup>9</sup>. The description of the results is facilitated by the fact that the dielectric constant,  $D$ , in all dilute solutions, is a linear function of the concentration,  $C$ , of amino acid in moles per liter.

$$D = D_0 + \delta C \quad (6)$$

where  $D_0$  denotes the dielectric constant of the pure solvent. For most amino acids, indeed, this linear relation holds up to the highest concentrations investigated. The dielectric increment  $\delta = dD/dC$  is characteristic of the substance, and related to its polarity. Values for  $\delta$  in water for many substances are listed in Tables 2 and 3.

Table 2 shows that almost all molecules (excluding dipolar ions), even highly polar ones such as acetone, acetonitrile and acetamide, give negative  $\delta$  values in water. In a homologous series, such as alcohols, all of which have the same dipole moment, the value of  $\delta$  decreases progressively as the number of  $-\text{CH}_2$  groups in the molecule increases. This may be attributed to a "dilution" of the effect of the dipoles by the non-polar hydrocarbon groups. The number of dipoles in a mole of liquid remains constant, but the number per unit volume (which is the significant quantity in determining the dielectric constant) decreases<sup>10</sup>.

For molecules of somewhat similar size, Table 3 shows that the value of  $\delta$  tends to increase with the dipole moment. Of all the molecules listed (excepting dipolar ions), only malonamide, urea, thiourea, and *sym*-dimethylurea show positive  $\delta$  values. Since urea, with a moment of 8.6, has a dielectric increment of about 3, it may be inferred that glycine, with a dielectric increment of 22.6, must possess a moment very considerably greater than that of urea or thiourea.

Furthermore, the dielectric increment  $\delta$  for amino acids is relatively independent of solvent and temperature. The effect on  $\delta$  of variation in the solvent is shown for  $\alpha$ -aminobutyric acid in Table 4 taken from Wyman<sup>9</sup>. In general, the value of  $\delta$  tends to decrease as the dielectric constant of the solvent increases, but the variation from solvent to solvent is relatively small. However, betaine and the benzbetaines, which are soluble in solvents of very low dielectric constant, show a very marked fall

TABLE 3. DIELECTRIC INCREMENTS OF AMINO ACIDS AND OTHER DIPOLAR IONS IN WATER

(Unless otherwise indicated, the $\delta$ values are at 25°)		
Substance	Ref.	Dielectric Increment, $\delta$
(a) <i>Amino Acids</i>		
Glycine.....	(1, 2, 3, 4)	22.6; 23.0; 26.4; 30
$\alpha$ -Alanine.....	(1, 2, 3)	23.2; 23.6; 27.7
$\beta$ -Alanine.....	(1)	23.2
$\gamma$ -Aminobutyric acid.....	(1)	22.6
$\delta$ -Aminovaleric acid.....	(5)	25
<i>dl</i> -Valine.....	(5)	25
<i>l</i> -Leucine.....	(5)	21
<i>dl</i> -Proline.....	(5)	24.5
Sarcosine.....	(5)	26
<i>d</i> -Glutamic acid.....	(3)	27.8
<i>l</i> -Aspartic acid.....	(6)	20.8
<i>d</i> -Glutamine.....	(3, 6)	28.4; 20.4
<i>l</i> -Asparagine.....	(7)	ca. 30
<i>N</i> -Phenylglycine.....	(6)	30
Glycocylamine.....	(6)	32.2
Creatine.....	(1, 8, 2)	34.6; 35; 42.3
$\beta$ -Alanine.....	(1, 9)	32.4; 36
$\beta$ -Aminobutyric acid.....	(5)	41
Taurine.....	(8)	51
$\gamma$ -Aminobutyric acid.....	(1)	54.8
$\gamma$ -Aminovaleric acid.....	(9)	51
Ornithine.....	(8)	63
$\delta$ -Aminovaleric acid.....	(10)	62
Arginine.....	(6)	62
Acetylhistidine.....	(1, 8)	77.5; 73
$\epsilon$ -Aminocaproic acid.....	(8)	87
$\zeta$ -Aminoheptylic acid.....	(b) <i>Peptides</i>	
Diglycine.....	(1, 5, 11, 4)	70.6; 70; 70.5; 80
Glycylalanine.....	(6)	71.8
Alanylglycine.....	(6)	71
Leucylglycine.....	(11, 6)	62; 68.4
Glycylleucine.....	(6, 11)	74.6; 70
<i>N</i> -Methylleucylglycine.....	(6)	67
Glycylphenylalanine.....	(6)	70.4
Phenylalanylglycine.....	(6)	56.7
Triglycine.....	(1, 11)	113; 128
Leucylglycylglycine.....	(6, 11)	120.4; 112
Tetraglycine.....	(1)	159
Pentaglycine.....	(1)	215
Hexaglycine.....	(1)	234
Heptaglycine.....	(12)	290 (in 5.14 molar urea)
(c) <i>Double Dipoles</i>		
$\epsilon, \epsilon'$ -Diguanido - di ( $\alpha$ - thio - <i>n</i> - caproic acid).....	(6)	151

TABLE 3—Continued

Substance	Ref.	Dielectric Increment, $\delta$
(d) <i>Aliphatic Betaines</i>		
Betaine ( <i>N</i> -trimethyl glycine) . . . . .	(9, 15)	24-27; 18.2
Betaine of $\alpha$ -amino- <i>n</i> -valeric acid . . . . .	(4)	60
Betaine of $\zeta$ -amino-pentadecyclic acid . . . . .	(4)	220 (at 70°)
Betaine of $\pi$ -amino-heptadecyclic acid . . . . .	(4)	190 (at 80°)
(e) <i>Aromatic Betaines</i>		
Pyridine betaine . . . . .	(15, 4)	18.5; 20.5
Ortho-benzbetaine . . . . .	(15, 7)	18.7; 20
Meta-benzbetaine . . . . .	(15, 7)	48.4; 58
Para-benzbetaine . . . . .	(15, 7, 4)	72.4; 68; 62

The order in which the  $\delta$ -values for any substance are listed corresponds to the order of the references from which the values were obtained.

$\delta$ -values for aminobenzoic acids and alicyclic amino acids are tabulated by J. Wyman, Jr., *Chem. Rev.*, **19**, 218 (1936); see also Chapter 5, Table 6, and footnote 31 on p. 154.  $\delta$ -values for many betaines and some other substances not listed here are also given in Wyman's review.

- (1) Wyman, J., Jr., *J. Am. Chem. Soc.*, **55**, 408 (1933).
- (2) Hedestrand, G., *Z. physiol. Chem.*, **18**, 36 (1928).
- (3) Devoto, G., *Gazz. Chim. Ital.*, **64**, 371 (1934).
- (4) Hausser, I., *Sitz. Akad. d. Wiss.*, No. 6 (1935).
- (5) Devoto, G., *Gazz. Chim. Ital.*, **63**, 50 (1933).
- (6) Greenstein, J. P., and Wyman, J., Jr., *J. Am. Chem. Soc.*, **58**, 463 (1936).
- (7) Devoto, G., *Gazz. Chim. Ital.*, **64**, 371 (1934).
- (8) Devoto, G., *Gazz. Chim. Ital.*, **63**, 50 (1933).
- (9) Devoto, G., *Gazz. Chim. Ital.*, **64**, 76 (1934).
- (10) Devoto, G., *Z. physiol. Chem.*, **222**, 227 (1933).
- (11) Cavallaro, L., *Arch. Sci. Biol. (Italy)*, **20**, 587 (1934).
- (12) Wyman, J., Jr., *J. Am. Chem. Soc.*, **56**, 536 (1934).
- (13) Greenstein, J. P., Wyman, J., Jr., and Cohn, E. J., *J. Am. Chem. Soc.*, **57**, 637 (1935).
- (14) Greenstein, J. P., Klemperer, F. W. and Wyman, J., Jr., *J. Biol. Chem.*, **125**, 515 (1938); **129**, 681 (1939).
- (15) Edsall, J. T., and Wyman, J., Jr., *J. Am. Chem. Soc.*, **57**, 1964 (1935).

or  $20^{11}$ . This effect may be due to association of the dipolar ions in such solvents, in such a way as to reduce the moment of the associated complex; or alternatively to a distortion of the valence bonds within the individual molecules, so as to bring the charged groups closer together.

Wyman and McMeekin<sup>12</sup> showed a small effect of temperature on  $\delta$  for glycine, which was 23.90 at 0° and 22.58 at 25°. This effect has recently been studied in more detail by Lindquist and Schmidt<sup>13</sup>; some examples of the results obtained are given in Table 5.  $\delta$  always increases somewhat with decreasing temperature, but the variation with temperature is found to be relatively small, though certainly significant.

The variation of the dielectric increment of glycine with pH has been studied by Dunning and Shutt<sup>13a</sup>, who found, as would be expected from its dissociation constants (Chapter 4), that  $\delta$  is constant over a considerable pH range (4.5 to 7.5) and falls steeply on either side beyond this range (Fig. 1). They attempted to calculate the dielectric constant of the solution from the known values of  $pK_1$  and  $pK_2$  of glycine, assuming that only

the dipolar ions of glycine present at any pH contribute to the observed dielectric increment, the anions and cations being regarded as equivalent to an equal volume of water. The values so calculated are shown by the dotted lines in Fig. 1. They lie distinctly above the observed values, and indicate that the assumptions made are inexact, although they give a good first approximation to the observed phenomena.

Urea, in contrast to glycine, gives a dielectric increment independent of pH over the entire pH range from 2 to 11. This was to be expected, as urea behaves neither as an acid nor as a base within this pH range.

The value of  $\delta$  is nearly the same for all  $\alpha$ -amino acids; but it increases rapidly with increase in the number of carbon atoms separating the amino and carboxyl groups (Fig. 2a). The increase in  $\delta$  is about 13 units for

TABLE 4. DIELECTRIC INCREMENT OF SOLUTIONS OF  $\alpha$ -AMINOBUTYRIC ACID (25°C.)

Solvent	$D_0^*$	$\delta$
80% ethyl alcohol.....	35.37	24.0
60% ethyl alcohol.....	47.20	22.1
40% ethyl alcohol.....	59.41	22.6
20% ethyl alcohol.....	69.82	23.6
Water.....	78.54	23.5
2.58 <i>M</i> urea.....	87.37	20.9
0.961 <i>M</i> glycine.....	99.8	23.1
1.993 <i>M</i> glycine.....	124.05	19.0
2.510 <i>M</i> glycine.....	134.9	18.4

\* The dielectric constant of the solvent.  
(From Wyman, J., Jr., *Chem. Rev.*, 19, 228 (1936)).

TABLE 5. EFFECT OF TEMPERATURE ON THE DIELECTRIC INCREMENT,  $\delta$ , IN CERTAIN AMINO ACIDS

Amino Acid	1°	Temperature in °C		
		10°	20°	30°
Glycine.....	26.03	25.23	23.36	22.33
<i>dl</i> -Alanine.....	26.43	25.61	24.05	22.86
<i>l</i> -Hydroxyproline.....	23.01*	21.86	20.83	20.16
<i>dl</i> -Proline.....	23.36*	—	—	21.47

\* Temperature 1.9°.

Data of Lindquist, F. E., and Schmidt, C. L. A., *Compt. rend. trav. lab. Carlsberg*, 22, 307 (1938).

each additional intervening carbon atom. The esters of the amino acids, which, of course, are not dipolar ions, show no indication of such behavior. The dipole moments of esters of  $\alpha$  and  $\beta$  amino acids<sup>14</sup> are identical practically within the limits of experimental error at about 2.1 Debye units. Similarly for the peptides of glycine (Fig. 2b), there is a practically linear relation between  $\delta$  and the number of glycy residues in the molecule, the increase for each additional residue being about 45. Glycine heptapeptide gives a dielectric increment of 290, and the double dipole, lysylglutamic acid, with two free ammonium and two carboxyl groups, gives an even higher value, 345<sup>15</sup>.

### The Dipole Moments of Dipolar Ions

Qualitatively, it is clear that high  $\delta$  values correspond to high values of the dipole moment,  $\mu$ ; but the exact nature of the functional relation between the two quantities is still not completely clarified. The dipole moment of an  $\alpha$ -amino acid, such as glycine, however, may be approximately estimated by considering a molecular model, based on x-ray diffraction studies<sup>16</sup>.

This model is pictured and described in detail in Chapter 14. If we take the center of positive charge as located on the nitrogen atom, and the center of negative charge midway between the two oxygens of the car-

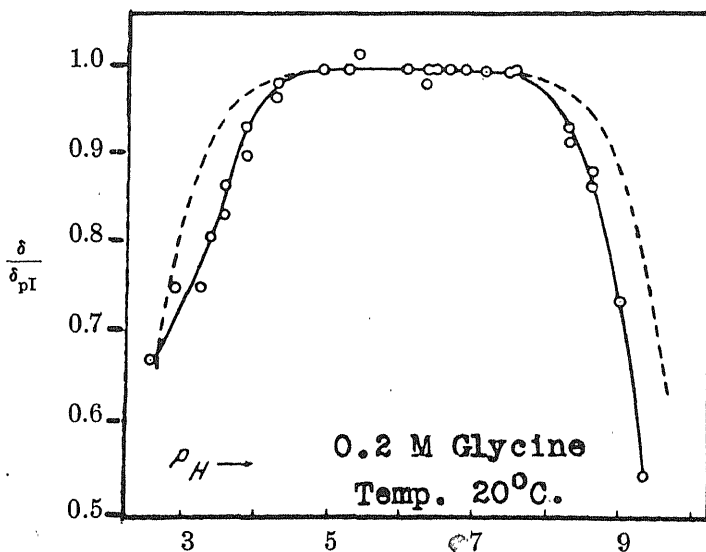


Figure 1. Variation of low-frequency dielectric constant of glycine with pH. Data of Dunning and Shutt.

boxylate group, the dipole distance,  $R$ , is found to be 2.96 Å, corresponding to a moment of 14.2 Debye units.

Two effects might tend to reduce the value of the moment so calculated: (1) Bending of the valence angle to bring the charged groups closer together. This effect is probably small, since the intramolecular forces resisting such a deformation are apparently very large<sup>17</sup>. (2) Production of secondary induced moments, of opposite sign to the principal moment determined by dipolar ionic structure. This effect appears also to be small since the moment<sup>18</sup> has been calculated from the effect of salts on the solu-

bility of glycine in alcohol-water mixtures, that the moment of the glycine molecule is of the order of 15 (see Chapters 11 and 12).

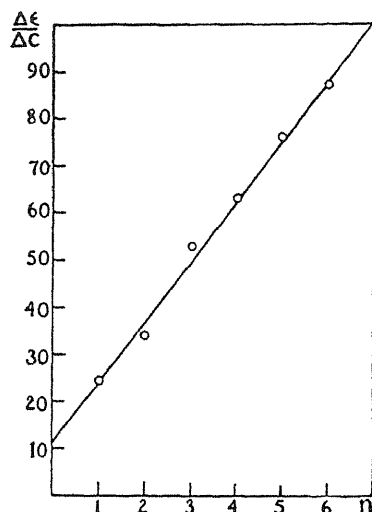


Figure 2a.

The dielectric increments of amino acids;  $n$  = number of carbon atoms between the amino and carboxyl groups. From Wyman, J., Jr., *Chem. Rev.*, 19, 213 (1936).

The symbol  $\Delta\epsilon/\Delta C$  is employed by Wyman to denote the increment called  $\delta$  in this chapter.

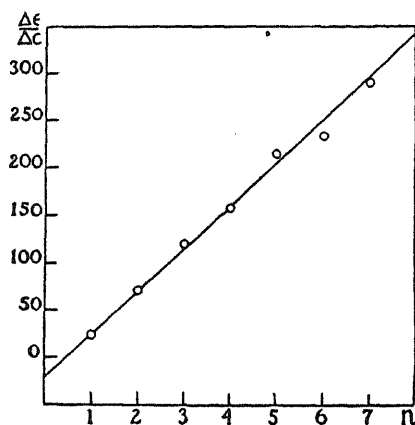


Figure 2b.

The dielectric increments of peptides of glycine;  $n$  = number of glycine units in the molecule. From Wyman, J., Jr., *Chem. Rev.*, 19, 213 (1936).

A further analysis of the dielectric constant data themselves by Wyman<sup>9, 19</sup> has led to similar conclusions. The relation between the dielectric constant and the polarization, which is expressed in equation (5), is valid

When applied to polar liquids, however, it leads to physically very improbable relations. In polar liquids, it seems more probable, both from experimental data<sup>20</sup> and from theoretical considerations<sup>21</sup> that there is approximately a linear relation between the dielectric constant and the polarization of the liquid per unit volume.

$$D = j + hp \quad (7)$$

where  $p$  is the polarization of the liquid per cubic centimeter and  $j$  and  $h$  are constants. The treatment of Wyman (20) and Onsager (21) indicates that, in pure polar liquids composed of molecules of known moment,  $j = -1$  and  $h$  lies between 6 and 10, the best value being approximately 8.5.

To obtain a simple treatment for dipolar ion solutions, we shall assume that the polarizations ( $p_s$  and  $p_0$ ) and partial molal volumes ( $V_s$  and  $V_0$ ) of solute and solvent are independent of concentration<sup>22</sup>. Thus for the pure solvent, from equation 7

$$D_0 = j + hp_0 \quad (7a)$$

If the concentration of solute is  $n_s$ , in moles per cc., then the volume fraction of solute is  $n_s V_s$ , and of solvent is  $1 - n_s V_s$ . Then the polarization per cc. of solution is

$$p = p_0(1 - n_s V_s) + p_s n_s V_s \quad (8)$$

From 7, 7a and 8:

$$D - D_0 = h(p - p_0) = hn_s V_s (p_s - p_0) = hn_s V_s \left( p_s - \frac{D_0 - j}{h} \right) \quad (9)$$

If we write  $C_s = 1000n_s$  for the concentration in moles per liter, and  $P_s = p_s V_s$  for the molal polarization of the solute, this becomes

$$D - D_0 = [hP_s - V_s(D_0 - j)] \frac{C_s}{1000} \quad (9a)$$

and for the dielectric increment

$$\delta = \frac{dD}{dC_s} = \frac{hP_s - V_s(D_0 - j)}{1000} \quad (10)$$

For dipolar ions, the term  $V_s(D_0 - j)/1000$  is small in comparison with the observed  $\delta$  value, and the dielectric increment is determined primarily by the molal polarization of the solute,  $P_s$ . This is in turn connected with the dipole moment,  $\mu$ , by the relation (see equation 5):

$$\mu^2 = \frac{9kT}{4\pi N} P_s, \quad \text{or} \quad \mu = 0.0127 \times 10^{-18} \sqrt{P_s T} \text{ esu.} \quad (11)$$

<sup>20</sup> Wyman, J., Jr., *J. Am. Chem. Soc.*, **58**, 1482 (1936).

<sup>21</sup> Onsager, L., *ibid.* **58**, 1488 (1936). See also Van Vleck, J. H., *J. Chem. Physics*, **5**, 556 (1937), and Kirk-



where we have neglected the optical polarization  $P_0 = (4\pi N/3)\alpha_0$ , arising from the polarizability, since for dipolar ions it is extremely small compared to the polarization arising from dipole orientation. Thus, according to this treatment, the dielectric increment of a dipolar ion is approximately proportional to the *square* of the dipole moment of the solute.

From equations 10 and 11 Wyman<sup>9</sup> has calculated the moments of several dipolar ions (in Debye units) as follows: glycine, 12.2;  $\beta$ -alanine, 15;  $\gamma$ -aminobutyric acid, 18;  $\delta$ -aminovaleric acid, 20;  $\epsilon$ -aminocaproic acid, 22; glycine tetrapeptide, 31; glycine hexapeptide, 38 (see also Chapter 12).

A convenient approximate empirical equation has also been used by Cohn,<sup>10</sup> p. 105. Since there is a linear relation between  $\delta$  and  $\mu^2$ ,

$$\delta = A\mu^2 - K \quad (12)$$

where  $A$  and  $K$  are constants.  $K$  is small and may tentatively be set equal to zero, and  $A$  can then be fixed by taking a probable value for the moment of glycine, say 15.2 Debye units (Kirkwood's estimate). This, corresponds to a distance,  $R$ , between the positive and negative charges of 3.17 Å, from which we obtain

$$R = \sqrt{\frac{\delta}{2.3}} \quad \text{or} \quad \mu = \sqrt{10\delta} \quad (13)$$

This simple relation is often very convenient for making approximate estimates of dipole moments of dipolar ions. (See also pp. 294-6).

#### The Problem of Free Rotation around Valence Bonds in Dipolar Ions

It is apparent from these data that the dipole moments of amino acids and peptides in solution do not increase proportionately with the number of atoms in the chain separating the charged amino and carboxyl groups. This indicates that the molecules in solution do not exist in the form of an extended zigzag chain, such as is found for the hydrocarbon chain in crystals of such molecules as the fatty acids. On the other hand, the charged groups in the peptides and long chain  $\omega$ -amino acids clearly do not approach as closely as they might; for in any long-chain compound, such as  $\epsilon$ -aminocaproic acid or glycylglycine, it would apparently be easy for the molecule to assume the form of a ring in which the amino and carboxyl groups approached each other as closely as their own dimensions would permit. The fact that the dipole moment increases with increasing number of atoms in the chain separating the charged groups shows that this ring configuration cannot be the predominant one. On statistical grounds, Werner Kuhn<sup>23</sup> has estimated the average distance of separation between

rotation around all valence bonds in the chain, and that there is no systematic relation between the angles of rotation around the successive bonds. On these assumptions, he finds the mean square of the distance between the ends of the chain,  $\bar{C}^2$ , to be proportional to the number,  $n$ , of the atoms in the chain.

$$\bar{C}^2 = nC_1^2 \left( \frac{1 + \cos \theta}{1 - \cos \theta} \right) \quad (14)$$

where  $\theta$  is the supplement of the valence angle,  $C_1$  is the distance between the centers of adjacent atoms of the chain, and the length of the chain is  $nC_1$ . Thus, if the valence angle is the tetrahedral angle, then  $\cos \theta = .333$ , and

$$\bar{C}^2 = 2nC_1^2 \quad (14a)$$

If this theory is correct, therefore, the dipole moment should be proportional to  $(\bar{C}^2)^{1/2}$ , which from 14 is proportional to  $\sqrt{n}$ . This result is in excellent accord with the dielectric constant measurements and the theoretical conclusions which have been drawn from them, since  $\delta$  is found in practice to be proportional to  $n$ , and should also be proportional to  $\mu^2$ , which likewise, from Werner Kuhn's theory, is proportional to  $n$ .

This treatment by Kuhn neglects two factors. First, it takes no account of the space occupied by the atoms in the chain. If we allow for this and make some estimate of the distance of closest approach between the atoms, this would cut down considerably the amount of free rotation which is estimated to be possible. On the other hand, the electrostatic attraction between the amino and carboxyl groups tends to draw them closer together. Thus these two effects which have been neglected work in opposite directions and at least partly cancel one another. In a later paper<sup>24</sup>, Kuhn has considered the influence of electrostatic forces on the mean distance of separation of the charged groups; owing to certain approximations introduced in his treatment, it is probable that the values he has so introduced for the dipole moments are too low. In any case, this paper furnished an interesting and suggestive treatment of an important problem.

A very elegant formula for the mean length of a chain of  $n$  atoms, assuming free rotation around valence bonds, has been given by Eyring<sup>25</sup>

$$\bar{C}^2 = C_1^2 [n + 2(n-1) \cos \theta + 2(n-2) \cos^2 \theta + \cdots + 2 \cos^{n-1} \theta] \quad (15)$$

where the symbols have the meaning already given. This formula, while more complex than equation 14, has the advantage of holding for all values of  $n$ , even for  $n = 1$  and  $n = 2$ , whereas 14 is valid only when  $n$  is

large<sup>26</sup> Eyring's formula has been applied by Wyman<sup>27</sup> to dipolar ions composed of a chain of atoms to which adjoining charged groups are attached at various points along the chain. Wyman thus calculated the mean square value of the dipole moment of such a structure, assuming free rotation. This moment was then related to the expected dielectric increment of such a molecule in water, by the treatment already outlined (equations 7 to 11, inclusive). For lysylglutamic acid, the dielectric increment calculated by this procedure was 323, the observed value of 345 being only 6% higher. Wyman concluded "that there is the same degree of free rotation in lysyl-glutamic acid as in the other simple peptides and amino acids which have been studied"<sup>27</sup> (p. 149).

Similarly, the dielectric increments of the alicyclic amino acids<sup>28</sup> and of the tetrapolar peptides cystinyldiglycine, cystinyldidiglycine, and diglycylcystine<sup>29</sup> have been shown to be such as would be expected on the hypothesis of free rotation around single bonds. It should be pointed out, however, that this does not necessarily imply free rotation in the strict sense. If the directed valence bonds formed by two linked carbon atoms can assume a few different relative orientations about the C—C link, taken as axis of rotation, and if all these orientations are about equally probable<sup>30</sup>, then the average state of a molecule containing many atoms in the chain is very nearly the same as if rotation were actually free. In view of recent work on hindrance to free rotation about single bonds, this latter interpretation of the data appears the most probable.<sup>31</sup>

<sup>26</sup> Both formulas, however, neglect the effects of steric hindrance.

<sup>27</sup> Wyman, J., Jr., *J. Physical Chem.*, **43**, 143, 285 (1939).

<sup>28</sup> Greenstein, J. P., and Wyman, J., Jr., *J. Am. Chem. Soc.*, **60**, 2341 (1938).

<sup>29</sup> Greenstein, J. P., Klemperer, F. W., and Wyman, J., Jr., *J. Biol. Chem.*, **125**, 515 (1938); **129**, 681 (1939).

<sup>30</sup> In ethane, for instance, there are three stable configurations for the relative orientations of the hydrogens about the C—C bond. Rotation is hindered by the energy barrier to passage from one stable configuration to another involves the surmounting of this barrier, which resists rotation. See Conn, J. B., Kistiakowsky, G. B., and Smith, E., *J. Am. Chem. Soc.*, **61**, 1613 (1939), and references cited there.

<sup>31</sup> Since this chapter was written, Conner, W. P., Clarke, R. P., and Smyth, C. P., *J. Am. Chem. Soc.*, **64**, 1379 (1942) have published dielectric increments of glycine and alanine, of the series of glycine peptides including hexaglycine, and of several peptides containing glycyl, alanyl and leucyl residues. Their measurements, carried out at 10 meters, are generally in excellent agreement with the data summarized in Table 3. Their  $\delta$ -values for tetraglycine (165.8) and pentaglycine (202.2) differ from those of Wyman and McMeekin. These authors also give a very careful and critical discussion of the relation between  $\delta$  values and dipole moments, with conclusions not dissimilar to those of this chapter and of Chapters 12 and 22. Further work of Conner and Smyth on relaxation times of amino acids and peptides is discussed in Chapter 22.

## Chapter 7

# Apparent Molal Volume, Heat Capacity, Compressibility and Surface Tension of Dipolar Ions in Solutions

BY JOHN T. EDSALL

It has long been known that the physical properties of aqueous solutions of strong electrolytes are unusual in several respects. If a salt is dissolved in water, the volume of the salt solution is always less than the sum of the volumes of the pure water and the solid salt of which it is composed. In some instances—for example, in dilute solutions of magnesium sulfate—the volume of the solution is actually less than that of the water in the system, without salt. Likewise, the heat capacity of dilute solutions of inorganic strong electrolytes is always less than that of the water in the system alone; and the compressibilities of such solutions are markedly less than that of the solvent water.

Many years ago, a hypothesis was advanced by Tammann<sup>1</sup> to explain these phenomena. He pointed out that the effects of strong electrolytes and other solutes on the physical properties of water were also produced by increase of pressure. Hence he inferred the existence of high internal pressures in such solutions, produced by the attraction of the water molecules around the ions or molecules of the solute. Tammann's theory assumed that not only ions, but also uncharged solute molecules, could exert such attractive forces in aqueous solution; and he advanced considerable evidence to support this conclusion. It was apparent, however, that free ions, because of the great electrostatic forces surrounding them, should increase the internal pressure far more than uncharged molecules. In the following year, this problem was further examined by Drude and Nernst<sup>2</sup> who pointed out that a process such as the ionization of water involves a contraction amounting to approximately 21 cc per mole of water ionized. Contractions of the same order of magnitude accompany the ionization of organic acids and bases. Drude and Nernst attributed these

hood of the ion, and falls off rapidly with increasing distance. Electric dipoles, such as water molecules, experience no translations, but only rotations, from the influence of a *homogeneous* electric field. In an *inhomogeneous* field, however, dipoles tend to move toward the region in which the field strength is maximal. Thus water molecules tend to be clustered and compressed in the neighborhood of an ion, whether the charge on the ion is positive or negative. (The orientation of dipoles around an ion naturally depends on the sign of its charge.)

Many of the properties of ionic solutions are in large measure explained by this electrostriction. Clearly dipolar ions also must produce electrostriction of the solvent. Dipolar ions, however, also contain various uncharged organic residues which must interact with the solvent in a different fashion. The electrical effects produced by a dipolar ion may be most clearly distinguished by comparing the properties of its solutions with those of an uncharged isomer containing polar groups of a similar nature. Thus to the simplest known dipolar ion, glycine,  ${}^+\text{H}_3\text{N}\cdot\text{CH}_2\cdot\text{COO}^-$ , there corresponds the uncharged isomer glycolamide  $\text{HO}\cdot\text{CH}_2\cdot\text{CO}\cdot\text{NH}_2$ . The differences between the properties of solutions of these two molecules reveal the distinctive effects of the charged groups in glycine; and similar comparisons may be made between larger dipolar ions and their isomers.

The quantitative description of the properties of solutions with which we are concerned in this chapter may be given most simply in terms of *apparent molal properties*, denoted by the symbol  $\Phi(G)$  and defined by the relation

$$\Phi(G) = \frac{G - n_1 G_1^0}{n_2} \quad (1)$$

Here  $G$  denotes any extensive property of the solution (such as volume, energy, free energy, heat capacity, entropy, etc.),  $G_1^0$  is the corresponding molal property of the solvent, and  $n_1$  and  $n_2$  denote the moles of solvent and solute respectively. Thus in a volume,  $V$ , of solution which contains *one mole* of solute, the apparent molal volume of the solute is equal to  $V$  minus the volume  $V_1$ , of the water used in making up the solution:

$$\Phi(V_2) = V - V_1 \quad (2)$$

It is clear from this equation that the error in  $\Phi(V_2)$  becomes larger as  $V$  increases, being nearly inversely proportional to the concentration  $C$ . The same is true of any apparent molal property; so that measurements of great precision are necessary to determine  $\Phi$  with accuracy in dilute solutions<sup>3</sup>

## APPARENT MOLAL VOLUMES

The apparent molal volume of a solute is most conveniently calculated from the density of the solution, by the equation

$$\Phi(V_2) = \frac{1000}{C} \left( \frac{d_1 - d}{d_1} \right) + \frac{M_2}{d_1} \quad (3)$$

where  $C$  is the concentration in moles per liter, and  $d$  and  $d_1$  are the density of the solution and of pure water.

A systematic study of the apparent molal volumes of many organic molecules in dilute aqueous solution was first undertaken by Traube<sup>4</sup>. He showed that two molecules differing only by a certain group—for instance, a  $\text{CH}_2$  group—differ also by a certain increment in volume corresponding to that group; and these increments could be expressed as the sum of volumes assigned to the individual atoms in the group. The molal volume increments for certain atoms and groups, deduced by Traube for a temperature of  $15^\circ\text{C}$ , are as follows:

	C	H	N	O'		O''
Volume of atoms, cc/mole	9.9	3.1	1.5	0.4	2.3*	5.5
	— $\text{NH}_2$		— $\text{CH}_2$	— $\text{COOH}$		— $\text{CONH—}$
Volume of groups, cc/mole	7.7		16.1	18.9		20.0

\* Traube believed that a second hydroxyl group, or the second oxygen in a carboxyl group occupied less space than the first.

The apparent molal volume of a compound, however, is always greater than the sum of the volumes assigned to its constituent atoms or groups. This additional volume—termed the “co-volume” by Traube—varies very little from molecule to molecule, and is approximately 13 cc per gram mole. The existence of this covolume must be associated with the repulsive intermolecular forces which set lower limits to the distances of approach between molecules in the liquid; but the nature of these forces need not concern us here.

The effect of temperature on these apparent molal volumes is relatively slight; Traube calculated that the volume of the  $-\text{CH}_2$  group varies from 16.0 cc at  $0^\circ\text{C}$ . to 16.3 cc at  $30^\circ\text{C}$ .

## Electrostriction Due to Charged Groups in Ions and Dipolar Ions

The amino acids form a striking exception to the general rules deduced by Traube. This deviation is most clearly revealed by a comparison of the amino acids with uncharged isomers which should have the same apparent

A series of such comparisons is given in Table 1.

The values of  $\Phi(V)$  are in all cases lower for the dipolar ions than for the isomeric uncharged molecules. Moreover, this lowering ( $E$ , in Table 1) is greater for  $\beta$ - than for  $\alpha$ -alanine, and decidedly greater for glycylglycine than for glycine. Since the number of polar groups is the same in the dipolar ions as in their uncharged isomers, the difference between them must be due to electrostriction arising from the charged groups. The results show that the electrostriction,  $E$ , increases with the distance between the charges.

Similar figures are found for the volume changes accompanying the ionization of weak electrolytes. Thus Weber<sup>5</sup>, in an important contribu-

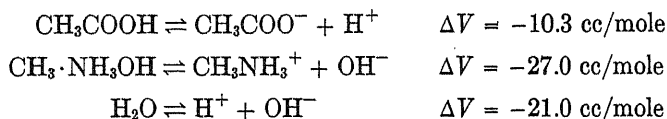
TABLE 1. APPARENT MOLAL VOLUMES  $\Phi$  OF DIPOLAR IONS AND THEIR UNCHARGED ISOMERS

( $E$  denotes the difference in each case between charged and uncharged isomers)

Substance	Ref.	$\Phi$ cc per mole	$E$ cc per mole
Glycolamide ( $\text{CH}_2\text{OH}\cdot\text{CONH}_2$ ).....	*	56.2	
Glycine ( $^+\text{H}_3\text{N}\cdot\text{CH}_2\cdot\text{COO}^-$ ).....	b	43.5	12.7
Lactamide ( $\text{CH}_3\cdot\text{CHOH}\cdot\text{CONH}_2$ ).....	*	73.8	
$\alpha$ -Alanine ( $^+\text{H}_3\text{N}\cdot\text{CH}(\text{CH}_3)\cdot\text{COO}^-$ ).....	b	60.6	13.2
$\beta$ -Alanine ( $^+\text{H}_3\text{N}\cdot\text{CH}_2\cdot\text{CH}_2\cdot\text{COO}^-$ ).....	b	58.9	14.9
Methylhydantoic Acid ( $\text{H}_2\text{N}\cdot\text{CONH}\cdot\text{CH}(\text{CH}_3)\cdot\text{COOH}$ ).....	c	94.2	
Glycylglycine ( $^+\text{H}_3\text{N}\cdot\text{CH}_2\cdot\text{CONH}\cdot\text{CH}_2\cdot\text{COO}^-$ ).....	b	77.2	17.0

\* J. Am. Chem. Soc., 56, 784 (1934).  
 b J. Am. Chem. Soc., 57, 626 (1935).  
 c J. Am. Chem. Soc., 57, 626 (1935).

tion to the study of dipolar ions, found the following values for the molal contraction accompanying ionization in solution.



In most cases, no uncharged compounds have been studied which permit a direct comparison with the amino acids as in Table 1. However, the apparent molal volume of such uncharged isomers may be calculated, either by the method of Traube<sup>4</sup>, or by the similar method of Cohn, McMeekin, Edsall, and Blanchard<sup>6</sup>. When ring structures are present, as in proline, oxyproline, tyrosine, tryptophane and phenylalanine, special corrections must be employed for the effect of the ring on the volume of the molecule. These "ring factors" have been discussed in detail by Traube.

Table 2 lists the observed apparent molal volumes for all amino acids hitherto studied, and for some related substances. Values of the electrostriction,  $E$ , are also tabulated. The electrostriction produced by the  $\alpha$ -amino acids ranges between 11.5 and 14.7 cc, averaging about 13.3 cc

TABLE 2. APPARENT MOLAL VOLUMES ( $\Phi$ ) AND ELECTROSTRICTIONS ( $E$ ) OF AMINO ACIDS, PEPTIDES AND RELATED COMPOUNDS IN WATER

Substance	$\Phi$ cc per mole	$E$ cc per mole	Substance	$\Phi$ cc per mole	$E$ cc per mole
(a) Amino Acids			(b) Peptides		
Glycine <sup>b</sup> .....	43.5	13.5	Diglycine <sup>b</sup> .....	77.2	16.1
$\alpha$ -Alanine <sup>b</sup> .....	60.6	12.7	Glycylleucine <sup>d</sup> .....	139.8	18.7
Serine <sup>b</sup> .....	60.8	12.9	Leucylglycine <sup>d</sup> .....	143.2	15.3
$\alpha$ -Amino- <i>n</i> -butyric acid <sup>b</sup> .....	76.5	13.1	Glycylalanine <sup>d</sup> .....	93.9	15.7
$\alpha$ -Aminoisobutyric acid <sup>b</sup> .....	78.1	11.5	Alanylglycine <sup>d</sup> .....	94.5	14.1
$\alpha$ -Amino- <i>n</i> -valeric acid <sup>b</sup> .....	92.7	13.2	Glycylphenylalanine <sup>d</sup> .....	155.4	15.9
$\alpha$ -Aminoisovaleric acid <sup>b</sup> .....	91.3	14.6	Phenylalanylglycine <sup>d</sup> .....	160	11.3
$\alpha$ -Amino- <i>n</i> -caproic acid <sup>b</sup> .....	108.4	13.8	Triglycine <sup>b</sup> .....	113.5	16.1
$\alpha$ -Aminoisocaproic acid <sup>b</sup> .....	107.5	14.7	Leucylglycylglycine <sup>d</sup> .....	178.5	16.3
<i>L</i> -Asparagine <sup>d</sup> .....	78.0	15.3	(c) Double Dipoles		
$\beta$ -Alanine <sup>b</sup> .....	58.9	14.4	$\epsilon, \epsilon'$ -diamino-di-( $\alpha$ -thio- <i>n</i> -caproic acid) <sup>e</sup> .....	226.8	27.1
$\beta$ -Aminobutyric acid <sup>b</sup> .....	76.4	13.2	Lysylglutamic acid <sup>e</sup> .....	172.7	38.8
$\gamma$ -Aminovaleric acid <sup>b</sup> .....	90.0	15.9	(d) Aromatic Amino Acids and Betaines		
$\epsilon$ -Aminocaproic acid <sup>b</sup> .....	104.9	17.3	<i>o</i> -Aminobenzoic acid <sup>b</sup> .....	96.7	5.9
Lysine <sup>b</sup> .....	108.5	18.3	<i>m</i> -Aminobenzoic acid <sup>b</sup> .....	90.3	12.3
Proline <sup>b</sup> .....	81.0	13.8	<i>p</i> -Aminobenzoic acid <sup>b</sup> .....	97.3	5.3
$\Pi$ -Hydroxyproline <sup>b</sup> .....	84.4	12.7	Betaine <sup>*f</sup> .....	97.7	8.2
Phenylalanine <sup>b</sup> .....	121.3	13.7	<i>o</i> -Benzbetaine <sup>*f</sup> .....	146.5	4.9
Histidine <sup>b</sup> .....	99.3	10.8	<i>m</i> -Benzbetaine <sup>*f</sup> .....	145.0	6.4
Acetylhistidine <sup>d</sup> .....	134.0	11	<i>p</i> -Benzbetaine <sup>*f</sup> .....	141.3	10.1
Tryptophane <sup>b</sup> .....	144.1	15.1	<i>N</i> -dimethylphenylglycine <sup>*f</sup> .....	146.8	4.6

\* The electrostriction produced by betaines is much greater in ethanol and ethanol-water and ethanol-benzene mixtures than in water.

Apparent molal volumes for uncharged organic molecules were calculated by the method of Traube<sup>4</sup> or by a modification of that method proposed by McMeekin, Edsall and Blanchard<sup>5</sup>. The difference between these calculated values and the observed values of  $\Phi$  gives the estimated electrostriction per mole. The values given in this table differ slightly from those given in the literature because of the use of the methods of calculation employed here. The values of  $E$  are calculated from the values of  $\Phi$  and may be taken as the probable values.

Some very careful measurements of the electrostriction of amino acid solutions have been reported by Dalton and his co-workers<sup>6</sup>. The values of  $E$  for the amino acids are given in Table 2. The electrostriction values for  $\beta$ -alanine and  $\beta$ -aminobutyric acid are 14.4 and 13.2 cc per mole, respectively, which are in good agreement with the values given in this table.

<sup>b</sup> Same as Reference 1. <sup>c</sup> Same as Reference 1. <sup>d</sup> Same as Reference 1. <sup>e</sup> Same as Reference 1. <sup>f</sup> Same as Reference 1.

The betaines, with three organic groups on the charged nitrogen atom, give much lower values—presumably because the bulky methyl or phenyl



values than any of the  $\alpha$ -acids, because of the greater separation of the charged groups. The same is true of all the peptides. As the charged groups become more and more widely separated, the electrostriction appears to approach asymptotically a certain limiting value, in the neighborhood of 20 cc/mole for compounds containing one  $-\text{NH}_3^+$  and one  $-\text{COO}^-$  group. In lysylglutamic acid, with two positive and two negative groups, widely separated, the electrostriction is nearly twice as great (38.8 cc/mole).

The values for the aminobenzoic acids indicate that the *meta* acid exists in water principally as a dipolar ion, the *ortho* and *para* acids chiefly as uncharged molecules. This is in accord with the conclusions reached from the study of dissociation constants (Chapter 5) and dielectric constants (Chapter 6).

Considering only spherical ions of a given charge, small ions should give rise to more electrostriction than larger ions; since the electric field strength near the ion, and the resulting compression of the dipoles of the solvent, are greater the smaller the radius within which the charge of the ion is included. A detailed calculation of the electrostriction produced by ions of different radii was made by Webb<sup>7</sup>, in the course of an investigation of the free energy of hydration of ions. The values obtained by Webb are certainly of the right order of magnitude, but he was forced in his calculations to make certain simplifying assumptions so that the exact magnitudes of the values obtained by him cannot be safely assumed as a basis for further calculation.

More recently, Bernal and Fowler<sup>8</sup>, in their important study of the properties of water, have developed a more detailed picture of the electrostriction process. The structure of liquid water, near the melting point, may be regarded as a partially broken down ice structure. In ice, each oxygen atom is surrounded tetrahedrally by four other oxygens, the O—O distance being 2.76 Å. The hydrogens lie approximately on lines connecting the oxygen atoms; two of the four hydrogens surrounding any one oxygen are chemically bonded to it, the O—H distance being 0.99 Å; the other two are connected to it by "hydrogen bonds", the O—H distance being 1.77 Å<sup>8a</sup>. The structure of water near the melting point is very similar to this, but more disarranged; a considerable number of the hydrogen bonds are broken, the number increasing with rising temperature.

The structure outlined above is a very loose, open one; each oxygen has only four nearest neighbors, whereas in a close packed liquid structure there would be twelve. This open structure is maintained primarily by electrostatic forces between the water dipoles. When an ion is introduced into

## DIPOLAR IONS IN SOLUTION

around the ion. Hence there is a shrinkage in volume, which Bernal and Fowler<sup>8</sup> have estimated by comparing the volume of a hydrated ion in a crystal, surrounded by  $n$  water molecules, (subtracting the volume of the same ion in an anhydrous crystal) with the volume of  $n$  water molecules in liquid water. They concluded that most of the electrostriction effect can be explained by hydration of ions in solution, to the same extent as in crystals. "For univalent ions of radius greater than 1.6 Å no hydration is to be expected, while for all monatomic polyvalent ions, hydration will always occur" (8, p. 533). This treatment considers only the electrostriction due to the packing of water molecules in the shell immediately surrounding the ion; furthermore, polyatomic ions, even when polyvalent, are assumed to be unhydrated. It is probable, however, that such ions as  $\text{SO}_4^{2-}$  and  $\text{CO}_3^{2-}$  produce considerable electrostriction, as judged by their marked salting-out effect (Chapter 11). Bernal and Fowler's treatment is thus only a first approximation, but as such it is extremely valuable.

The electrostriction produced by several salts (LiCl, NaCl, KCl, NaI, KI) in some aliphatic alcohols has been investigated by Vosburgh, Connell and Butler<sup>9</sup> who have shown that all these salts have much lower apparent molal volumes in alcoholic solution than in water. They have applied Webb's method of calculation to evaluate theoretically the electrostriction produced by these salts in alcohol, which they calculate (in agreement with experiment) to be from two to five times as great as in water, depending on the radius of the ion.

One group of dipolar ions, the betaines, are soluble in alcohol, and in alcohol-ether and alcohol-benzene mixtures. The electrostriction produced by several betaines in water and organic solvents has been investigated by Edsall and Wyman<sup>10</sup>; they find that for the benzbetaines (N-trimethylaminobenzoic acids) the electrostriction is four or five times as great in alcohol as in water, increasing for *p*-benzbetaine for instance from 10 cc/mole in water to 40 cc/mole in ethyl alcohol. These figures were shown to agree well with the values calculated by Vosburgh, Connell and Butler for relative electrostriction values in these two solvents.

### Apparent Molal Volume as a Function of Concentration

The apparent molal volumes of most substances vary with the concentration of the solution. In aqueous solutions  $\Phi(V)$  almost always increases with increasing  $C$ . The form of its functional dependence on  $C$ , however, reveals a fundamental difference between electrolytes and non-electrolytes.

many strong electrolytes could be represented as linear in the square root of the concentration

$$\Phi = \Phi_0 + q'C^{1/2} \quad (4)$$

This relationship was found to hold over a very wide concentration range—up to several moles per liter in some cases. On the other hand, Redlich and Rosenfeld<sup>13</sup> derived from the Debye-Hückel theory by thermodynamic reasoning a formula formally identical with equation 4, valid only in very dilute solutions.

According to the Debye-Hückel theory (see Chapter 3) the free energy,  $F$ , of a strong electrolyte solution should be given, in the limit of low concentration, by

$$F = n_1 F_1^0 + n_2 [\bar{F}_2^0 + \nu RT \ln m - Q(D, T) \omega^{3/2} C^{1/2}] \quad (5)$$

Here  $F_1^0$  and  $\bar{F}_2^0$  are the partial molal free energies of solvent and solute, respectively, in the reference states;  $D$  is the dielectric constant of the solvent;  $m$  the molality;  $Q$  is the function of  $D$  and  $T$  given by the Debye-Hückel theory (Chap. 3, page 54 ff);  $\omega = 1/2 \sum_i \nu_i z_i^2$ ;  $\nu_i$  is the number of ions of species  $i$  formed by one molecule of the electrolyte;  $\nu = \sum \nu_i$ ; and  $z_i$  = valence. Applying a fundamental thermodynamic relation<sup>14</sup> we obtain, on differentiating  $F$  with respect to the pressure,

$$\left( \frac{\partial F}{\partial P} \right)_{T,m} = V = n_1 V_1^0 + n_2 \left[ \bar{V}_2^0 + q \left( D, T, \frac{\partial D}{\partial P}, \beta \right) \omega^{3/2} C^{1/2} \right] \quad (6)$$

where  $\beta$  is the compressibility of the solvent. The function  $q$  is given by

$$q = R \left( \frac{2\epsilon^2}{Dk} \right)^{3/2} \left( \frac{\pi N}{1000T} \right)^{1/2} \left( \frac{1}{D} \frac{\partial D}{\partial P} - \frac{\beta}{3} \right) \quad (7)$$

Here  $k$  is Boltzmann's constant, and the pressure is in atmospheres.

Equation 6 may be written

$$\Phi = \Phi_0 + q\omega^{3/2}C^{1/2} \quad (6a)$$

which has the same form as 4, but a different meaning, since it demands that the slope of the  $\Phi$  curve, plotted against  $C^{1/2}$ , be the same for all electrolytes of a given valence type, and that it should be much higher for electrolytes of higher valence types than for 1-1 electrolytes. In contrast to this, Masson and Geffcken were forced to assign quite different coefficients,  $q'$ , to different salts of the same valence types, in order that they might fit equation 4, over a wide concentration range. Equation 6, however, is applicable only in very dilute solutions, and measurements made at high concentrations are irrelevant in testing its validity.

solutions of certain uni-univalent salts, Redlich<sup>15</sup> evaluates the coefficient,  $q$ , in equation 6a, as

$$q = 1.86 \pm 0.02 \text{ (25°C.)} \quad (8)$$

The theoretical value of  $q$  from 6 is more uncertain, chiefly because of uncertainties regarding the value of the coefficient  $\frac{\partial D}{\partial P}$  for water; but the data available are compatible with equation 8<sup>16</sup>. In any case, there appears little doubt of the validity of the square root law for the apparent molal volumes of electrolytes in dilute solution.

TABLE 3

1. APPARENT MOLAL VOLUME  $\Phi(V)$  OF UREA, SUCROSE AND CERTAIN AMINO ACIDS AT 25°, AS DESCRIBED BY THE EQUATION  $\Phi = \Phi_0 + aC + bC^2$

Substance	$\Phi_0$ (cc.)	$a$	$b$
Urea.....	44.218	0.13999	-.002601
Sucrose (15°).....	209.741	0.46185	.001412
Glycolamide.....	56.156	0.1595	0
Glycine.....	43.199	0.8614	0
$\alpha$ -Alanine.....	60.6 (60.609)	0.60 (0.5731)	0
$\beta$ -Alanine.....	58.6 (58.730)	0.72 (0.7117)	0
Lactamide.....	73.508	0.0169	0
$\alpha$ -Aminobutyric acid.....	76.5 $\pm$ 0.3	0	0
<i>dl</i> -Valine.....	91.05 $\pm$ 0.1	0	0
$\epsilon$ -Aminocaproic acid.....	104.35	0.72	0
Glycylglycine.....	76.8	1.30	0

2. Apparent Molal Volume of Amino Acid Salts, as Described By the Equation

$$\Phi = \Phi_0 + A\sqrt{C}$$

Substance	$\Phi_0$	$A$	Substance	$\Phi_0$	$A$
Na Glycinate.....	42.4	3.89	Na $\epsilon$ -Aminocaproate..	106.0	
Glycine HCl.....	67.0	2.54	$\epsilon$ -Aminocaproic HCl..	135.4	

The values for urea, sucrose,  $\alpha$  and  $\beta$ -alanine and lactamide, are the most accurate. Other substances, except *dl*-valine, were studied only at concentrations of 0.25 molar and above.

Urea values from F. T. Gucker, Jr., E. W. Gage and C. E. Moser, *J. Am. Chem. Soc.*, 60, 2582 (1938); sucrose—see O. Redlich and H. Krigbaum, *Trans. Faraday Soc.*, 65, 1155 (1969); *dl*-Valine—see J. B. Dalton and C. E. Moser, *J. Am. Chem. Soc.*, 60, 1611 (1938);  $\alpha$ -alanine—see F. T. Gucker, Jr., W. L. Ford and C. E. Moser, *J. Am. Chem. Soc.*, 60, 1611 (1938);  $\beta$ -alanine—see F. T. Gucker, Jr., W. L. Ford and C. E. Moser, *J. Am. Chem. Soc.*, 60, 1611 (1938);  $\epsilon$ -aminocaproic acid—see F. T. Gucker, Jr., W. L. Ford and C. E. Moser, *J. Am. Chem. Soc.*, 60, 1611 (1938); glycylglycine—see F. T. Gucker, Jr., W. L. Ford and C. E. Moser, *J. Am. Chem. Soc.*, 60, 1611 (1938).

Values for lactamide, and values in parentheses for  $\alpha$ - and  $\beta$ -alanine, from T. W. Allen, *Thesis*, Northwestern Univ. (1941).

At one time it was suggested<sup>3</sup> that a law of the same form might hold for non-electrolytes. Careful studies on sucrose<sup>17</sup> and on urea<sup>18</sup>, however, showed that the apparent molal volumes of these substances in dilute solution were linear functions of the first power of the concentration with no square root term. The studies of Daniel and Cohn<sup>19</sup> led to the same conclusion, for certain isoelectric amino acids (Table 3 and Fig. 1). Thus molecules carrying zero net charge, even if they are dipolar ions, obey a

<sup>15</sup> O. Redlich, *J. Physical Chem.*, 44, 619 (1940).

<sup>16</sup> The findings in this field have given rise to considerable controversy. The situation up to 1933 is discussed by

different law from that holding for electrolytes. The slope of the  $\Phi$ - $C$  curve, however, denoted by  $\alpha$  in Table 3, is much greater for most of the amino acids than for urea and sucrose, which are not dipolar ions. For the amino acids,  $\alpha$  is greater, the shorter the hydrocarbon chain and the greater

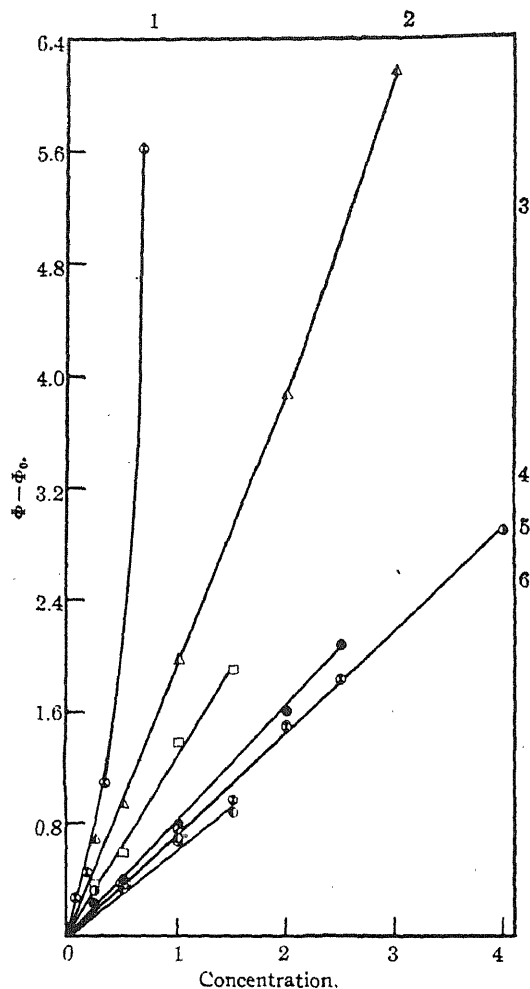


Figure 1.

Increase in apparent molal volume as a function of the concentration: (1) lysyl-glutamic acid; (2) sodium  $\epsilon$ -aminocaproic acid; (3) glycylglycine; (4) glycine; (5)  $\epsilon$ -aminocaproic acid, and  $\beta$ -alanine; (6)  $\alpha$ -alanine. From Daniel, J., and Cohn, E. J., *J. Am. Chem. Soc.*, 58, 415 (1936).

the dipole moment. For the aminobutyric acids  $\Phi$  does not increase with concentration; and the studies of Dalton and Schmidt<sup>20</sup> have shown that

methionine. Lysylglutamic acid (Fig. 1) ( $\Phi_0 = 172.4$ ) shows a very steep  $\Phi-C$  curve, with a slope of between 2 and 3 in dilute solution; the curve rises much too rapidly to be described by a simple linear equation, however. This behavior is probably associated with its very high dipole moment.

The sodium salt and the hydrochloride of glycine, in contrast to glycine itself, appear to obey the square root law, like simple strong electrolytes (Table 3, Part 2) at least as a first approximation. On the other hand, the sodium salt and the hydrochloride of  $\epsilon$ -aminocaproic acid do not obey the square root law, the curves not being linear either with  $C$  or with  $C^{1/2}$ . It would appear that molecules containing a large number of non-polar groups do not obey the square root law, even if they are strong electrolytes. All these conclusions remain subject to revision. At present no highly accurate density measurements on very dilute solutions of dipolar ions are available; but it is undoubtedly in the region of high dilution, for dipolar ions as for electrolytes, that regularities of behavior should be most apparent. \* Investigations in this field remain to be reported. \*

#### APPARENT MOLAL HEAT CAPACITIES

The apparent molal heat capacity  $\Phi(C_p)$  of a solute is best evaluated from specific heat measurements, by the equation<sup>3</sup>

$$\Phi(C_p) = s \left[ \frac{1000}{m} + M_2 \right] - \frac{1000}{m} \quad (9)$$

where  $s$  is the specific heat of the solution,  $m$  is the molality, and  $M_2$  is the molecular weight of the solute.

Apparent molal heat capacities of non-electrolytes in water are almost always positive, and often not greatly different from the molal heat capacities of the same substances in the pure state. Strong electrolytes in water almost invariably have negative apparent molal heat capacities; that is, the heat capacity of the solution is less than that of the water in it alone. Some typical data are presented graphically in Fig. 2. These data are plotted against  $\sqrt{C}$ , since it was found empirically<sup>21</sup> by Randall and Rossini that  $\Phi(C_p)$  is a linear function of  $\sqrt{m}$  (or  $\sqrt{C}$ ) over a wide range. All the extrapolated values of  $\Phi$  at  $C = 0$  are negative, while the slope of the curves is positive. Salts of higher valence type ( $\text{CaCl}_2$  and  $\text{BaCl}_2$ ) have much lower  $\Phi$  values at  $C = 0$ , and much steeper slopes than the uni-univalent electrolytes.

An extensive theoretical treatment of the heat capacities of electrolyte solutions was given by Zwicky<sup>22</sup> who pointed out, on the basis of data

obtained by Bridgman<sup>23</sup>, that the heat capacity of water diminishes markedly with increasing pressure<sup>24</sup>. He therefore evaluated the pressure arising from electrostriction of the solvent, as a function of distance from

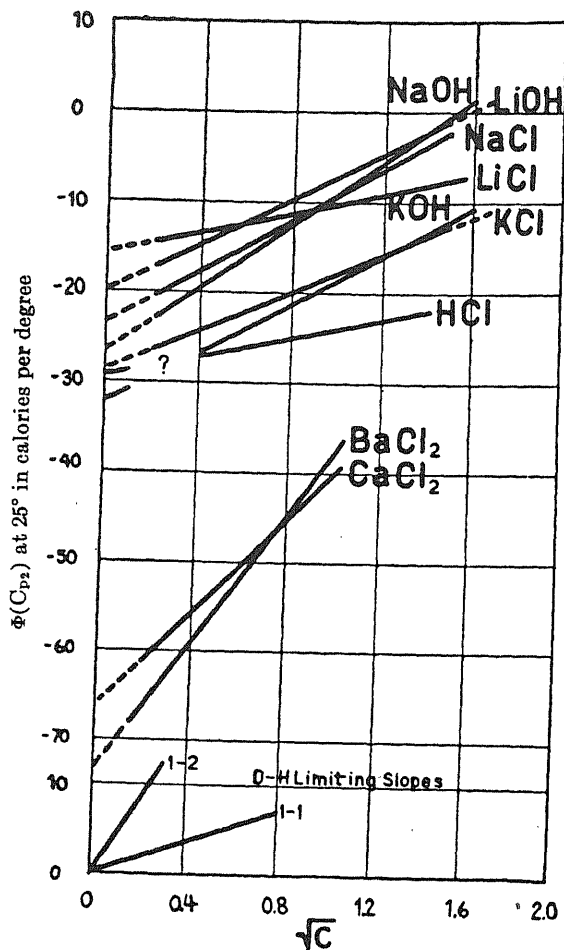


Figure 2.  
Apparent Molal heat capacities at 25°. From Gucker, F. T., Jr., *Chem. Rev.*, 13, 111 (1933).

the ion producing the electrostriction, and utilized the data of Bridgman to evaluate the effect of this pressure on the heat capacity of the water in the

<sup>23</sup> P. W. Bridgman, *Proc. Am. Acad. Arts & Sci.*, 48, 309 (1912-13).

<sup>24</sup> This holds for all pressures employed in Bridgman's investigations, at temperatures up to 60°C.; but at

solution. Other effects, which also were considered, involved the additional degrees of freedom in the system introduced by the presence of the solute molecules (giving a positive contribution to the heat capacity), and the polarization of water dipoles by the ions (giving a negative contribution). Both these effects, however, were small in comparison with the compression effect. Zwicky calculated the pressure due to electrostriction as of the order of 70,000 kg/cm<sup>2</sup> at 1 Å from the center of the ion; the pressure falls rapidly with increasing distance until it is only 23 kg/cm<sup>2</sup> at 10 Å from the center of the ion. Zwicky evaluated  $\Phi(C_p)$  for a 1-1 electrolyte as -117 cal/deg. Experimental values for electrolytes of this type, extrapolated to infinite dilution (Fig. 2) range from -15 to -32 cal/deg. Therefore the theory, while giving a qualitative picture of the underlying situation which is certainly valuable, is far from being quantitatively exact.

Zwicky's theory deals with infinitely dilute solutions and does not consider the interionic forces which cause  $\Phi(C_p)$  to vary with concentration. As with apparent molal volumes, however, the Debye-Hückel theory leads to a relation of the form

$$\Phi(C_p) = \Phi + \text{const. } C^{1/2} \quad (10)$$

for very dilute solutions. The constant in equation 10 is a function of the valence type of the electrolyte, the temperature, the dielectric constant of the solvent and its temperature derivatives, and the thermal expansion of the solvent<sup>25</sup>. The exact values of some of these quantities are not very accurately known, and some uncertainty therefore remains, as to the exact value of the slope to be expected for the  $\Phi-C^{1/2}$  curve. The experimental data in very dilute solutions are also somewhat uncertain<sup>26</sup>; but the general validity of the square root law for electrolytes appears to be almost universally recognized.

In contrast, non-electrolytes in dilute solution appear generally to obey a law of the form:

$$\Phi(C_p) = \Phi_0 + aC \quad (11a)$$

or

$$\Phi(C_p) = \Phi_0 + am \quad (11b)$$

analogous to the equation describing their apparent molal volumes. At higher concentrations, a term in  $C^2$  may be required to describe the data adequately. Probably the first study on a non-electrolyte to be carried out with very high precision was that of Gucker and Ayres<sup>27</sup> on sucrose.

<sup>25</sup> D. Zwicky, *Interpretation*, see V. K. La Mer and I. A. Coward, *J. Am. Chem. Soc.*, **55**, 1004 (1933).



Their results indicated that  $\Phi(C_p)$  for this substance in dilute solution varied with the first power of the molality; the equations describing the data were:

$$\Phi(C_p) = 145.87 + 1.950m - 0.1182m^2, \text{ at } 20^\circ$$

$$\Phi(C_p) = 151.20 + 1.325m - 0.0704m^2, \text{ at } 25^\circ$$

Neither in this nor in any other non-electrolyte, that has been accurately studied, has it proved to be necessary to introduce a square root term in describing the data.

The first measurements on the heat capacities of amino acid solutions were made by Zittle and Schmidt<sup>28</sup>, on glycine, *dl*-alanine and *dl*-valine at 25°. Specific heats were measured with an accuracy of 0.1%; and the results may be described, up to  $m = 1.5$ , by the equations:

$$\text{Glycine, } \Phi(C_p) = 7.5 + 1.0m \quad (12)$$

$$\text{dl-Alanine, } \Phi(C_p) = 40.0 - 3m \quad (13)$$

$$\text{dl-Valine, } \Phi(C_p) = 93 - 17m \quad (14)$$

These data reveal an extraordinary effect of the  $-\text{CH}_2$  group. The increment in  $\Phi$  for each additional  $\text{CH}_2$  is  $30 \pm 3$  cal/mole. In contrast, when homologous series of pure liquids are considered, the increment per  $\text{CH}_2$  group is only 6 to 8 cal/mole. A search of other data in the literature by Edsall<sup>29</sup> revealed similar very large increments in  $\Phi$  per  $\text{CH}_2$  group in the fatty acids and alcohols (Table 4). Thus the presence of a hydrocarbon chain tends to produce a *positive* deviation from the ideal solution laws for heat capacity. This trend is approximately indicated by the series of values of  $\Phi - C_p$  given in the last column of Table 4<sup>30</sup>. Whether this represents a general rule can only be decided when far more numerous and more accurate data are available. Substances rich in polar groups, such as glycerol, urea and dextrose (Table 4) behave much more nearly like ideal solutes in water, although sucrose shows a marked difference between  $\Phi$  and  $C_p$ .

The effect of dipolar ionization on  $\Phi(C_p)$ , like that of ionization, must clearly be to decrease it. Edsall<sup>29</sup> estimated the order of magnitude of the decrease to be expected, from EMF data on the ionization of fatty acids, ammonia and water, as follows<sup>31</sup>:

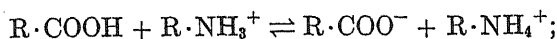
<sup>28</sup> C. A. Zittle and C. T. A. Schmidt, *J. Biol. Chem.*, **108**, 161 (1935).

<sup>29</sup> J. T. Edsall, *J. Am. Chem. Soc.*, **55**, 171 (1933).

<sup>30</sup> Preferably the values of  $\Phi$  and  $C_p$  should be those of  $\Phi$ , should be given for the pure liquid state. This is impossible for many substances, such as the amino acids which decompose on melting. In general, the value of  $C_p$  for any substance is distinctly higher for the liquid state than for the crystalline solid at or near the same temperature. This fact must be allowed for in considering the data of Table 4, but it does not



On adding 15a or 15b to 16 and 17, we obtain,



$$\Delta C_{p,ss} = -43 \text{ to } -47 \text{ cal/deg} \quad (18)$$

TABLE 4. APPARENT MOLAL HEAT CAPACITIES IN WATER.

Substance	Temp. (°C.)	N <sub>2</sub>	$\Phi$	$\Delta$	C <sub>p</sub>	$\Phi - C_p$
Formic Acid.....	16-50	0.02	17		24.6	-8
Acetic Acid.....	25	.02	38	21	32.2	6
Propionic Acid.....	22-50	.02	63	25	38.8	24
<i>n</i> -Butyric Acid.....	23-50	.02	85	22	44.5	40
Methyl Alcohol.....	20	.03	36		20.4	16
Ethyl Alcohol.....	20	.03	57	21	26.6	30
Propyl Alcohol.....	20	.015	88	31	34.2	54
<i>i</i> -Butyl Alcohol.....	25-29	.02	(160)		42.4	?
Glycine.....	25		8.0		24 <sup>a</sup>	-16
<i>dl</i> -Alanine.....	25	.01	38	30	29 <sup>a</sup>	9
<i>dl</i> -Valine.....	25	.01	85	24		
Glycerol.....	15	.02	59		51.1	8
Urea.....	16-20	.01	23		19.3 <sup>a</sup>	4
Dextrose.....	20	.005	60		54 <sup>a</sup>	6
Sucrose.....	20	.003	145		102.2 <sup>a</sup>	13

\* Indicates that substance in pure state is a crystalline solid.  
 $N_2$  represents the molecular fraction of solute at which the measurement was made;  $\Phi$  represents the heat capacity of the substances in the pure state at approximately the same temperature at which the solution was measured;  $\Phi - C_p$  represents the difference between the heat capacity and heat capacity in the pure state. The value of  $\Phi$  for alanine is to be regarded as very dubious.  
 The value of  $\Delta$  for alanine differs from alanine by two  $\text{CH}_2$  groups, and the given value of  $\Delta$  is estimated from the value for 7-oxo-2-oxoheptanoic acid, (not for comparison solutions from Gucker and Ayres, 1934; alanine from Huffman and Borst, 1935).

This should give the order of magnitude of the molal change in heat capacity accompanying the formation of a carboxylate and an ammonium ion, separated by a large distance, in dilute aqueous solution, whether the charges reside on the same or on different molecules. In  $\alpha$ -amino acids, owing to the close proximity of the charges, the effect should be less.

Essentially the same results may be derived from recent accurate EMF measurements on the ionization of  $\alpha$ -amino acids<sup>32</sup>, which permit the calculation of  $\Delta C_p$  for these ionization processes. In any case, the  $\Delta C_p$  data involve the second derivative of the EMF measurements, and are thus subject to some uncertainty.

$\Phi(C_p)$  was made by Gucker, Ford and Moser<sup>33</sup>, who studied glycine and its uncharged isomer glycolamide. Specific heats were measured with an average reproducibility of 0.01%, so that the results were far more accurate than those of the earlier amino acid investigations. The studies were carried out at three temperatures, 5°, 25° and 40°; the results are shown graphically in Fig. 3, and summarized in Table 5.

The decrease in  $\Phi(C_p)$  due to dipolar ionization (electrostriction) is 30.85 cal/deg at 5°, 26.93 at 25°, and 25.98 at 40°, which is close to what might be expected from the EMF measurements. It is notable that  $\Phi$  is negative

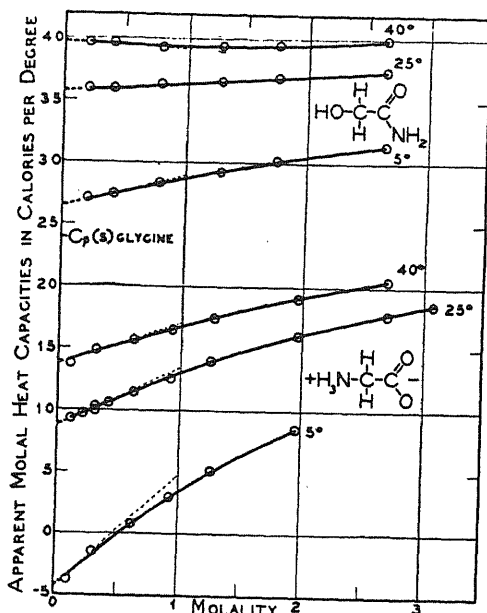


Figure 3.

Apparent molal heat capacities of glycine and glycolamide at 5°, 25°, and 40°. From Gucker, F. T., Jr., Ford, W. L., and Moser, C. E., *J. Physic. Chem.*, 43, 153 (1939).

for glycine at 5°; the only negative value of  $\Phi(C_p)$  for a non-electrolyte that has yet been reported.

The slopes,  $a$  (Table 5) of the  $\Phi - m$  curves in dilute solution are consistently much larger for glycine than for glycolamide at all temperatures. Indeed,  $a$  is negative for glycolamide at 40°; and for both substances its value decreases rapidly with increasing temperature. Fuoss<sup>34</sup> has calculated, from a general theory of dipole-dipole interaction, the apparent molal heat capacities of dipole solutes in dilute solution. In his theory the solute molecule is treated as a point dipole of moment  $\mu$ , located in the center of a sphere of radius  $a$ , and his analysis leads to the following

that  $\Phi(C_p)$  should be proportional to the first power of the concentration in dilute solution, and that the proportionality factor should increase with the dipole moment. Unfortunately, however, Fuoss' equations are not readily applied to substances of the dipole moment of glycine, in solvents of the dielectric constant of water; and no quantitative comparison between experiment and theory has yet been made.

The rapid variation of  $\Phi(C_p)$  with temperature (Table 5) is notable. For glycine at infinite dilution it increases by more than 18 cal/deg between 5° and 40°; for glycolamide by more than 13 cal/deg over the same range. In contrast, the molal heat capacity of crystalline glycine changes only by 5.1 cal/deg (from 18.9 to 24.0) in the 74 degree temperature interval from -47.5° to +26.4°<sup>35</sup>. A similarly great temperature

TABLE 5. APPARENT MOLAL HEAT CAPACITIES OF AMINO ACIDS AND THEIR UNCHARGED ISOMERS, AS DESCRIBED BY THE EQUATION  $\Phi(C_p) = \Phi_0 + am + bm^2$

Substance	Temperature 5°			Temperature 25°			Temperature 40°		
	$\Phi_0$	$a$	$b$	$\Phi_0$	$a$	$b$	$\Phi_0$	$a$	$b$
Glycine.....	-4.30	+9.01	-1.25	+8.83	+4.58	-0.47	+13.74	+3.27	-0.28
Glycolamide.....	+26.55	+2.39	-0.21	+35.76	+0.62	0	+39.72	-0.52	+0.23
$\alpha$ -Alanine.....	+23.75	+3.70	-0.20	+33.69	+1.48	+0.09	+36.66	+1.99	-0.22
$\beta$ -Alanine*.....	+4.04	+6.67	-0.86	+18.27	+2.64	-0.13	+23.34	+2.12	-0.10
Lactamide.....	+52.44	-0.43	+0.05	+58.38	-0.48	+0.02	+60.92	-0.57	+0.02

\* The complete equation for  $\beta$ -alanine (from  $m = 0$  to  $m = 4.53$ ) at 5° is

$$\Phi(C_p) = 4.04 + 6.67m - 0.86m^2 + 0.05m^3$$

All values in calories per degree per mole.

Data for glycine and glycolamide from F. T. Gucker, Jr., W. L. Ford and C. E. Moser, *J. Physical Chem.* **43**, 153 (1939); for  $\alpha$  and  $\beta$ -alanine and lactamide from T. W. Allen, *Thesis*, Northwestern University (1941). Gucker, F. T., Jr. and Allen, T. W. *J. Am. Chem. Soc.* **64**, 191 (1942).

Values of the partial molal heat capacity  $\bar{C}_{p_2} = \frac{\partial C_p}{\partial m}$  are given by the equation

$$\bar{C}_{p_2} = \Phi(C_p) + m \frac{\partial \Phi}{\partial m} = \Phi_0 + 2am + 3bm^2$$

dependence of  $\Phi(C_p)$  is found in urea<sup>36</sup>. Here  $\Phi$  varies from 10 cal/deg at 2° to nearly 25 cal/deg at 40°; while the molal heat capacity of crystalline urea varies only from 17.3 cal/deg at -50° to 22.4 at +25°<sup>35</sup>. This high temperature dependence of  $\Phi(C_p)$  in non-electrolytes is in marked contrast to the relative independence of temperature found for  $\Phi(V)$ . This suggests that a systematic attempt to correlate apparent molal heat capacity with structure may prove a more complex problem than in the case of apparent molal volumes.

Since this chapter was written, T. W. Allen in Gucker's laboratory has determined the apparent molal heat capacities of  $\alpha$  and  $\beta$  alanine and lactamide. His results are included in Table V, and plotted in Fig. 4.

The difference in  $\Phi(C_p)$  between  $\alpha$ -alanine and lactamide is nearly the same at any temperature as that between glycine and glycolamide.  $\Phi(C_p)$  for  $\beta$ -alanine is 20 units lower than for  $\alpha$ -alanine at 5°, and 13 units lower at 40°, showing the increase in electrostriction when the charged groups are

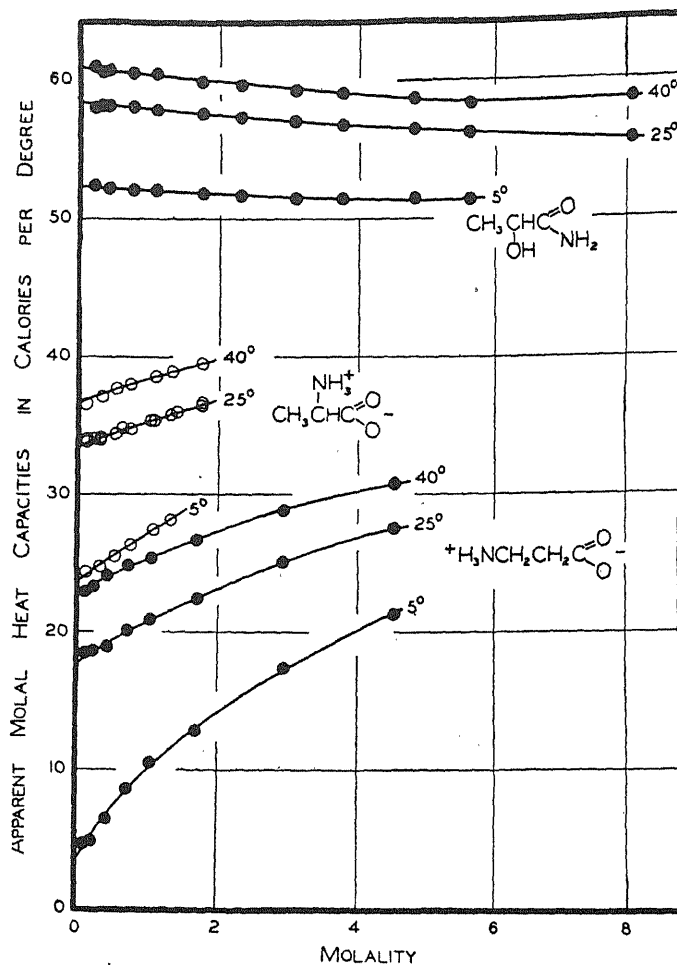


Figure 4. Apparent molal heat capacities of  $\alpha$ - and  $\beta$ -alanine and lactamide. From T. W. Allen, *Thesis*, Northwestern University, 1941.

more widely separated. The effect of the  $\text{CH}_2$  group on  $\Phi(C_p)$ —compare

positive in the order lactamide,  $\alpha$ -alanine,  $\beta$ -alanine—that is they increase with increasing dipole moment of the solute.<sup>36a</sup>

#### APPARENT MOLAL COMPRESSIBILITY

This function is defined by the equation:

$$\Phi(K_2) = \beta V - \beta_1 V_1 \quad (19)$$

or

$$\Phi(K_2) = \beta \left[ \frac{1000}{C} - \frac{\beta_1}{d_1} \frac{1000d}{C} \right] - M_2 \quad (20)$$

where the compressibility  $\beta = -\frac{1}{V} \left( \frac{\partial V}{\partial P} \right)_T$ , with a corresponding definition for  $\beta_1$ . ( $V$  and  $V_1$  have the same meaning as in equation 2.) The values of the apparent molal compressibility are negative for electrolyte solutions, and are a linear function of  $C^{1/2}$  over a wide range of concentration and temperature<sup>3</sup>. From the definition,

$$\Phi(K_2) = \frac{-\partial \Phi(V_2)}{\partial P} \quad (21)$$

and if  $\Phi(K_2)$  is negative, the apparent molal volume must *increase* with pressure (see reference 3).

The compressibility of solutions of glycine,  $\alpha$ -aminobutyric acid, and  $\epsilon$ -aminocaproic acid has been studied by Bridgman and Dow<sup>37</sup>. The results they obtained are exceedingly complex; it is difficult to make any generalization from their data, except that the apparent molal volumes generally *decrease* with increasing pressure—that is, the apparent molal compressibilities are *positive*. Acetic acid<sup>38</sup> also shows a positive value of  $\Phi(K_2)$ , but scarcely any other accurate data on organic solutes appear to be available.

Recently we have been informed by Dr. R. E. Gibson of the Carnegie Institution of Washington, D. C., that he has studied the apparent molal compressibility of glycine over a wide range of pressures, obtaining a much simpler type of curve than that found by Bridgman and Dow. Gibson's curve also shows a much closer similarity to that characteristic of electrolyte solutions than the data of Bridgman and Dow would indicate. Clearly the data in this field require reinvestigation, and the measurements should be extended to include a much wider variety of substances. For the present, attempts to correlate apparent molal compressibilities with the structure of organic compounds must be deferred.

## SURFACE TENSIONS OF AMINO ACID SOLUTIONS

The surface tension of aqueous solutions of practically all organic compounds is lower than that of water, although certain sugars form an exception to this rule<sup>39, 40</sup>. In any homologous series, the lowering of surface tension produced by a given concentration of solute increases in a regular way with the number of  $\text{CH}_2$  groups in the chain. Solutions of most

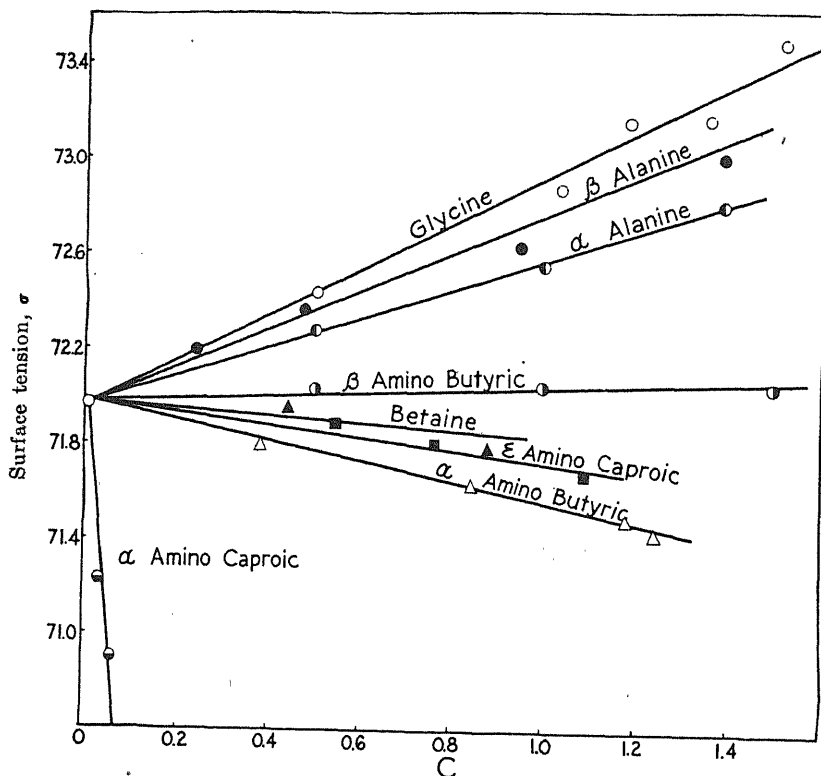


Figure 5. Surface tension of solutions of isoelectric amino acids at 25°. From Pappenheimer, J. R., Lepie, M. P., and Wyman, J., Jr., *J. Am. Chem. Soc.*, **58**, 1851 (1936).

inorganic strong electrolytes, however, show higher surface tensions than pure water. This means, of course, in accordance with Gibbs' theorem, that these ions, in contrast to the organic compounds, are negatively adsorbed at the liquid surface.

From their behavior in other respects, it might be

ions should show resemblances both to strong electrolytes and to typical organic compounds, with respect to their behavior at interfaces. The first systematic investigation of this point was that of Pappenheimer, Lepie and Wyman<sup>41</sup>, who determined the surface tension of a number of amino acids by the drop weight method. They found the relation between surface tension and concentration to be linear within the experimental error; the results may therefore be characterized by the slopes of the  $\sigma - C$  curves, which are given in Table 6 and represented graphically in Fig. 5. The following discussion is largely based on that given in their paper.

Recently glycine, alanine,  $\alpha$ -aminobutyric acid and valine have been studied by Belton<sup>42</sup>, who used a modified maximum bubble pressure method. His values of  $\Delta\sigma/\Delta m$  for the first three of these amino acids (glycine 1.06 to 1.15, alanine 0.86 to 1.06,  $\alpha$ -aminobutyric acid  $-0.20$  to  $-0.40$ ) are somewhat higher than those of Pappenheimer, Lepie and Wy-

TABLE 6. MOLAL SURFACE TENSION INCREMENTS  $\frac{\Delta\sigma}{\Delta C}$  OF ISOELECTRIC AMINO ACIDS AT 25°

Substance	$\Delta\sigma/\Delta C$
Glycine.....	+0.92
$\alpha$ -Alanine.....	+0.58
$\alpha$ -Aminobutyric Acid.....	-0.41
$\alpha$ -Aminocaproic Acid.....	-23.0
Betaine.....	-0.14
$\beta$ -Alanine.....	+0.77
$\beta$ -Aminobutyric Acid.....	+0.05
$\epsilon$ -Aminocaproic Acid.....	-0.25

From J. R. Pappenheimer, M. P. Lepie and J. Wyman, Jr., *J. Am. Chem. Soc.*, **58**, 1851 (1936).

man, and vary somewhat from one molality to another. Belton gives  $\Delta\sigma/\Delta m$  for valine as  $-4.36$ .

Glycine,  $\alpha$ -alanine,  $\beta$ -alanine and  $\beta$ -aminobutyric acid all behave like inorganic salts, in that the surface tension of their solutions is greater than that of water, although the molal surface tension increments are considerably less than for salts. Comparison of the four  $\alpha$ -amino acids, or the two  $\beta$ -amino acids, listed in Table 6, shows that increasing the length of the hydrocarbon chain, when the electric moment is held constant, decreases the molar increment of surface tension, as with other homologous series. The depressing effect on  $\Delta\sigma/\Delta C$  of each additional  $\text{CH}_2$  group increases approximately geometrically with the number of groups introduced. In this respect the amino acids behave similarly to other homologous series which have been studied.

A comparison of the isomers betaine and valine is of interest. Each contains three  $\text{CH}_2$  groups more than glycine, and their electric moments



are nearly identical, but in betaine the  $\text{CH}_2$  groups are attached to the charged nitrogen atom, while in valine they extend in a chain from the  $\alpha$ -carbon atom.  $\Delta\sigma/\Delta C$  for betaine is only  $-0.14$ , less negative than for  $\alpha$ -aminobutyric acid; while Belton's value of  $\Delta\sigma/\Delta m$  for valine ( $0.420m$ ) is  $-4.36$ . Thus the effect of a  $\text{CH}_2$  group on this increment depends very greatly on the place at which the group is attached, especially on whether or not it is part of a charged group. When isomers differing in dipole moment are compared, it is found that the isomer of greater moment gives always the more positive, or less negative, value of  $\Delta\sigma/\Delta C$ . Compare for instance, in Table 6 and Fig. 4,  $\alpha$ - and  $\beta$ -alanine,  $\alpha$ - and  $\beta$ -aminobutyric acid, and especially  $\alpha$ - and  $\epsilon$ -aminocaproic acid.

Pappenheimer, Lepie and Wyman determined the effect of pH on the surface tension of  $\epsilon$ -aminocaproic acid solutions, and found the tension to be a maximum at or near the isoelectric point. There was a marked drop in surface tension when the amino acid was converted either into the hydrochloride or the sodium salt. An effect of the same sort, although less marked, was found by Jones and Lewis<sup>43</sup> in glycine.

<sup>43</sup> R. Jones and W. C. M. Lewis, *Biochem. J.*, **26**, 638 (1932).

## Chapter 8

# Solubility of Amino Acids, Peptides and Related Substances in Water and Organic Solvents\*

BY JOHN T. EDSALL AND GEORGE SCATCHARD

Nowhere, perhaps, is the influence of dipolar ionic structure more apparent than in the solubility of dipolar ions. Here again, however, their distinctive behavior is most clearly appreciated when they are compared with other organic and inorganic compounds. Certain general relations between solubility and chemical structure have long been recognized. Substances which dissolve readily in water, for instance, are distinguished by the presence of certain polar groups, especially hydroxyl, carboxyl and amino groups; while the presence of a long hydrocarbon chain in the molecule tends to depress solubility in water, and to increase it in most organic solvents. Thus in a homologous series such as the alcohols or fatty acids, the lower members show a strong attraction for water; while the higher members are nearly insoluble in water but dissolve readily in ether, benzene or carbon tetrachloride<sup>1</sup>.

The relative solubility of an amino acid in two different solvents may often be predicted with considerable accuracy, and the systematic relations found are set forth in this and in the three following chapters. On the other hand, it is as yet impossible to predict the absolute value of the solubility of an amino acid in any one solvent, although important correlations between solubility and structure are observed. It is desirable, however, to consider certain general principles determining solubility, before passing on to the discussion of experimental data.

### Solubility and Intermolecular Forces

The solubility of a solid in a liquid may be conveniently factored into two parts: the solubility the solid would have in an ideal solution, and the activity coefficient in the solution in question. If the solid phase is the

\* Earlier reviews of much of the material in this chapter have been given by Cohn, E. J., *Ann. Rev. of Bio-*  
*chem.* 1939, 8, 450-480; and by Edsall, J. T., in Chapter XVI of "The Chemistry of the

pure solute, and not a solid solution, and if the heat of fusion may be regarded as independent of the temperature, we obtain from equation 42, Chapter 3,

$$-\ln a_2 = -\ln s_2 - \ln \gamma_2 = \frac{\Delta H_f/n}{RT_f} \left( \frac{T_f - T}{T} \right) \quad (1)$$

in which  $a_2$  is the activity of the solute in the saturated solution,  $s_2$  its mole fraction in the saturated solution, and  $\gamma_2$  its activity coefficient.  $\Delta H_f/n$  is the heat of fusion per mole, and  $T_f$  is the temperature of fusion.  $a_2$  may be regarded as the ratio of the vapor pressure of the solid to that of the supercooled liquid solute at the same temperature, assuming that the vapor behaves as a perfect gas.

In an ideal solution  $\ln \gamma_2$  is zero, so  $a_2 = s_2$ . This ideal solubility depends upon  $\frac{\Delta H_f}{nT_f}$ , and upon  $\frac{T_f}{T} - 1$ . For many substances it has been found that  $\frac{\Delta H_f}{nT_f}$  has approximately the same value, particularly when closely related substances are considered. Walden<sup>2a</sup> gave this value as 13.5 cal deg<sup>-1</sup> mole<sup>-1</sup> for many organic compounds. Where such a relation holds, it gives a law for the entropies of fusion corresponding to Trouton's law for the entropies of evaporation. Exceptions are found in the case of substances which undergo a transition in the solid state below the temperature of fusion; in this case the entropy of fusion may be abnormally low<sup>2b</sup>. Low entropies of fusion are found also for solids which melt to form "associated" liquids, especially those in which "hydrogen bonds" play an important part. Since amino acids decompose on melting, their entropies of fusion have not been determined, but it is certain that they have very high values, both of  $\Delta H_f/n$  and of  $T_f$ .

The ideal solubility also depends on  $\frac{T_f}{T} - 1$ . Thus at the melting point ( $T_f$ ),  $a_2 = 1$ , and above this point the liquid solute is of course miscible in all proportions with any liquid in which it forms an ideal solution. At temperatures considerably below  $T_f$ , equation 1 becomes inaccurate, since it neglects the change in  $\Delta H_f/n$  with temperature, which is determined by the difference in heat capacity between the solid solute and the pure liquid which it forms on melting. The equation can, however, be readily modified to take account of these terms (reference 1, pp. 32-35). The ideal solubility of any solid always increases with increasing temperature.

To consider  $\ln \gamma_2$  we investigate the work, other than the work of expansion, due to the forces of attraction. If the substance is very dilute

soluble this work is the same in the saturated solution as in the pure solvent. By definition it is the same in the ideal solution as in the supercooled liquid solute. So it is convenient to consider the transfer of molecules of solute from the supercooled liquid to an infinitely dilute solution in component 1 (the solvent). It is also convenient, in a preliminary discussion, to consider that the media are continuous, without molecular structure. Thus the surface of a hole left by the removal of a molecule is equal to the surface of the molecule.

The first step is to remove the molecule of surface area  $\mathcal{A}_2$  from the supercooled liquid into the vapor phase, leaving a hole in the interior of the liquid. If the work per unit surface is  $W_{22}$ , the work of this step is  $2\mathcal{A}_2W_{22}$ ; since two new surfaces—that of the molecule and that of the hole—have been created, and each is of area  $\mathcal{A}_2$ . When the hole in the liquid collapses, half this work is regained, making the net work in this process  $\mathcal{A}_2W_{22}$ , which is the energy of vaporization per molecule. Before the molecule can be transferred to the solvent, a hole must be created for it, which requires the work  $\mathcal{A}_2W_{11}$ . There is a gain in work on placing the molecule in the hole, for the surface of the molecule is covered by the solvent giving a work  $-\mathcal{A}_2W_{12}$ , and the surface of the hole is covered by the solute, giving a work  $-\mathcal{A}_2W_{21}$ . Since these operations cannot be carried out independently, it is not possible to determine  $W_{12}$  and  $W_{21}$  independently, so we usually let  $W_{12}$  represent the average and  $2\mathcal{A}_2W_{12}$  the work of this step. Then the total work is:

$$RT \ln \gamma_2(1) = \mathcal{A}_2(W_{11} + W_{22} - 2W_{12}) \quad (2)$$

The expression in parentheses is symmetrical with respect to the two components if the work depends only on the surfaces of contact. In this case

$$RT \ln \gamma_1(2) = \mathcal{A}_1(W_{11} + W_{22} - 2W_{12}) \quad (3)$$

The conceptions here outlined are due to Langmuir<sup>3</sup> and were elaborated by him in detail. Langmuir's treatment involved the assumption that the surface of a molecule is proportional to the two-thirds power of its volume. This is not in general true; and in most cases it appears more accurate to treat the work terms as proportional to the volume of the molecule<sup>4, 5</sup>. Thus the equations for  $\ln \gamma$  become

$$RT \ln \gamma_2(1) = V_2(b_{11} + b_{22} - 2b_{12}) \quad (4)$$

$$RT \ln \gamma_1(2) = V_1(b_{11} + b_{22} - 2b_{12}) \quad (5)$$

Scatchard<sup>4</sup> has shown that, for *non-polar liquids*,  $V_2b_{22}$  may be taken as the energy of evaporation per mole of the second component.

as that for the first component. If  $b_{12}$  is the arithmetic mean of  $b_{11}$  and  $b_{22}$ , the solutions formed are ideal. For non-polar mixtures in general, and often even when one component is somewhat polar,  $b_{12}$  is nearly equal to the geometric mean,  $\sqrt{b_{11}b_{22}}$ , so that  $(b_{11} + b_{22} - 2b_{12})$  is approximately  $(b_{11}^{1/2} - b_{22}^{1/2})^2$ , which is always positive.

If the molecules are distributed at random in the solution, both as regards position and orientation, then it is plausible to assume that the entropy of mixing is the same as for an ideal solution. If there is no change of volume on mixing<sup>6</sup>, then we may set  $\Delta H = \Delta E$  for the mixing process, and

$$\Delta F = \Delta H + N_1RT \ln N_1 + N_2RT \ln N_2 \quad (6)$$

Under these circumstances,  $b_{12}$  is independent of composition, and the activity coefficients in any solution of 1 and 2 are given by:

$$RT \ln \gamma_2 = V_2(b_{11} + b_{22} - 2b_{12}) \left( \frac{V_1 N_1}{V} \right)^2 \quad (7)$$

$$RT \ln \gamma_1 = V_1(b_{11} + b_{22} - 2b_{12}) \left( \frac{V_2 N_2}{V} \right)^2 \quad (8)$$

so that the treatment may be extended to solutions which are not dilute. If  $b_{12} = \sqrt{b_{11}b_{22}}$ , equations 7 and 8 become

$$RT \ln \gamma_2 = V_2(b_{11}^{1/2} - b_{22}^{1/2})^2 \left( \frac{V_1 N_1}{V} \right)^2 \quad (7a)$$

$$RT \ln \gamma_1 = V_1(b_{11}^{1/2} - b_{22}^{1/2})^2 \left( \frac{V_2 N_2}{V} \right)^2 \quad (8a)$$

It has been shown by Scatchard<sup>4</sup>, Hildebrand<sup>5</sup> and others that these equations describe satisfactorily the activity coefficients in many solutions of non-polar or slightly polar substances. These have been called "regular" solutions.

### Polar Liquids and Hydrogen Bonds

The situation in highly polar liquids is quite different, especially in liquids like water and the alcohols, in which neighboring molecules tend to assume preferred orientations relative to one another. The general treatment underlying equations 4 and 5 still applies, however, and in all cases  $\ln \gamma$ , as a function of composition, should be expressible in the power series discussed in Chapter 3.  $V_1 b_{11}$  and  $V_2 b_{22}$  can no longer be set equal to the energies of evaporation of components 1 and 2, but the energies of evaporation still afford a valuable index of the strength of the intermolecular forces in the pure liquids. Thus  $b_{11}$  in water is much greater than  $b_{22}$  in most

equation 4; hence the activity coefficient of the hydrocarbon is very large, and its solubility extremely low. Organic molecules are readily soluble in water only when they contain highly polar groups with a strong attraction for water. Dipolar solutes containing hydroxyl, carboxyl or amino groups are generally far more soluble in water than other molecules of the same or higher dipole moment (reference 1, p. 149). Thus propionitrile,  $C_2H_5CN$ , and nitroethane,  $C_2H_5NO_2$ , with dipole moments about  $3.5 \times 10^{-18}$  esu., are only moderately soluble in water, while  $C_2H_5OH$ , with a dipole moment only half as large, is miscible with water in all proportions. This effect of the OH dipole is due to the very small size of the proton, which permits it, when attached to a strongly electronegative atom, to attract a second electronegative atom to a very small distance of approach. This linkage of two electronegative atoms by hydrogen is known as the hydrogen bond<sup>7</sup>. It is relatively weak, compared with intramolecular bonds; the energy of the reaction,  $XH + Y = XH \cdots Y$ , being of the order of magnitude of 5 kcal/mole. This is sufficient, however, to give a very large positive contribution to  $b_{12}$  in equations 4 and 5. This tends to decrease  $\log \gamma$  and to increase the solubility of such a solute.

It is primarily the formation of hydrogen bonds between water molecules which makes  $b_{11}$  so large for water. In ice, which closely resembles cold water in structure, each hydrogen links the oxygens of two neighboring water molecules. The hydrogen is not, however, equidistant between the two; it is closely attached to one of them, within the molecule of which it forms a part (O-H distance about  $0.99\text{\AA}$ ), while it is more distant from the other (about  $1.77\text{\AA}$ ) (reference 7, p. 301), but is still much closer to it than would be possible without the strong electrostatic attraction which leads to hydrogen bond formation<sup>8</sup>.

It is primarily for steric reasons that the hydrogen bond is so important in enhancing cohesive forces between molecules. Other dipoles are limited to a much greater distance of closest approach because of the large radii of the atoms of which they are composed; hence the energy of interaction of the dipoles cannot be so great as in hydrogen bond formation, even though their moments may be larger.

Consider now the state of a molecule, say an alcohol, in dilute solution in water. The alcohol consists of a polar hydroxyl group attached to a non-polar hydrocarbon residue. As Langmuir<sup>9</sup> first showed in detail, it is often valuable to consider  $b_{12}$  for molecules containing several types of groups as made up of several terms. In the present case there would be two such terms, one representing the cohesive energy per unit surface (or volume) for the hydrocarbon-water interaction, the other representing

portional to the surface (or volume) of the groups involved. Extensive illustrations of the value of this concept are found later in this and the next chapter, in the discussion of relative solubilities in water and organic solvents. As a preliminary example, we may consider some data on dilute solutions of alcohols in water, since they reveal certain features characteristic of many other aqueous solutions (Table 1).

The values of  $\log \gamma$  in Table 1 were obtained by Butler<sup>9</sup> from measurements of the partial pressure,  $p_2$ , of alcohol vapor over dilute aqueous

TABLE 1. DILUTE SOLUTIONS OF ALCOHOLS IN WATER AT 25°

Alcohol	$\gamma = \frac{p_2}{N_2 p_2^0}$	$\log \gamma$	$\frac{RT}{\text{cal./mole}} \ln \gamma$	$\bar{H} - \bar{H}^0$ cal./mole	$\Delta S - \Delta S_I$	$\bar{C}_p - C_p^0$ cal./mole/deg.
CH <sub>3</sub> OH.....	1.51	0.179	244	-2300	-8.5	12
C <sub>2</sub> H <sub>5</sub> OH.....	3.69	0.567	774	-2760	-11.9	26
<i>n</i> -C <sub>3</sub> H <sub>7</sub> OH.....	14.4	1.158	1580	-3370	-16.6	51
<i>iso</i> -C <sub>3</sub> H <sub>7</sub> OH.....	7.7	0.887	1210	(-2830)	(-13.6)	
<i>n</i> -C <sub>4</sub> H <sub>9</sub> OH.....	52.9	1.724	2350	-4140	-21.7	
<i>iso</i> -C <sub>4</sub> H <sub>9</sub> OH.....	43.2	1.636	2230	(-2180)	(-14.8)	
<i>sec</i> -C <sub>4</sub> H <sub>9</sub> OH.....	25.1	1.400	1910			
<i>tert</i> -C <sub>4</sub> H <sub>9</sub> OH.....	11.8	1.072	1470	(-3550)	(-16.8)	
<i>n</i> -C <sub>5</sub> H <sub>11</sub> OH.....	214.	2.330	3180	-5050	-27.6	
<i>tert</i> -C <sub>5</sub> H <sub>11</sub> OH.....	35.	1.544	2110			
<i>n</i> -C <sub>6</sub> H <sub>13</sub> OH.....	903.	2.956	4040			
<i>n</i> -C <sub>7</sub> H <sub>15</sub> OH.....	3560.	3.551	4850			
<i>n</i> -C <sub>8</sub> H <sub>17</sub> OH.....	12300.	4.090	5580			

Value of  $\gamma$  from Butler, J. A. V., *Trans. Faraday Soc.*, 33, 229 (1937). Values for the heat of solution in water at infinite dilution,  $\bar{H} - \bar{H}^0$ , are taken partly from Butler, partly from "International Critical Tables" and Landolt-Börnstein "Ergänzungsbande," II and III.  $\Delta S - \Delta S_I$  represents the entropy of solution at infinite dilution, minus the entropy of solution for an ideal solution.

$$T(\Delta S - \Delta S_I) = \bar{H} - \bar{H}^0 - RT \ln \gamma$$

Values of  $\bar{C}_p - C_p^0$  calculated from Bose, E., *Z. Physik Chem.*, 58, 585 (1907). In all cases, the standard state is taken as the pure alcohol at 25°.

solutions. If  $p_2^0$  is the vapor pressure of the pure alcohol at 25°, and  $N_2$  the mole fraction of alcohol in the solution, then

$$\gamma_2 = \frac{p_2}{N_2 p_2^0}$$

This, extrapolated to zero mole fraction of alcohol, gives  $\gamma_2(1)$ . The reciprocal of  $\gamma_2(1)$  has been called the "ideal" solubility of component 2 in 1, since it is the mole fraction which would have unit activity, if the activity coefficient were the same in all solutions as in the infinitely dilute solution<sup>10</sup>.

<sup>9</sup> Butler, J. A. V. *Trans. Faraday Soc.* 33, 229 (1937).

The data in Table 1 show several characteristic features. The deviations from Raoult's law are invariably positive ( $\gamma > 1$ ). Considering only the straight-chain alcohols,  $\gamma$  increases between three- and fourfold for each additional  $\text{CH}_2$  group. A similar rule holds more exactly, as we shall see, for the relative solubility of amino acids and related substances in water and organic solvents. A branched-chain alcohol gives a lower value of  $\log \gamma$  than its straight-chain isomer; that is, it behaves like a straight-chain compound with fewer carbon atoms.

A great deviation from the laws of "regular" solutions is revealed by the comparison of the values of  $RT \ln \gamma$  and of  $\bar{H} - \bar{H}^0$  (heat of solution) in Table 1. For regular solutions—that is, solutions which fulfil the conditions assumed in deriving equation 7a—these two quantities should be identical (see equation 6). Here they are not only of very different magnitude, but of opposite sign. Hence there is a large, and negative, entropy of mixing, in addition to the entropy of mixing of an ideal solution. This shows again what we already know, that the positions and orientations of water molecules about an alcohol molecule are not random; although the magnitude and sign of this entropy change cannot as yet be predicted theoretically. We may infer from this and other facts that the values of  $b_{12}$  (and also of  $b_{11}$  and  $b_{22}$ ) for these systems must change considerably with temperature, and probably also with composition. The fact that heat is evolved on dissolving these alcohols in water shows that the activity coefficients increase with rising temperature (equation 46, chapter 3). In other words the positive deviations from Raoult's law increase with temperature. This phenomenon is characteristic of many organic solutes in water, which frequently show decreasing solubility with increasing temperature<sup>11</sup>; it is the converse of the usual behavior of non-aqueous solutions, which generally become more nearly ideal as the temperature rises (reference 1, pp. 49 and 57). The large positive values of  $\bar{C}_p - C_p^0$  (Table 1), however, tend to make  $\bar{H} - \bar{H}^0$  less negative at higher temperatures, so that at sufficiently high temperatures  $\frac{\partial \ln \gamma}{\partial T}$  may become negative.

Thus in many respects, the behavior of these solutions is complex. The change in free energy of solution, however, with change in the number of  $\text{CH}_2$  groups in the solute, is quite systematic and bears a definite relation to structure.

The substances considered in this preliminary discussion are dipoles of

---

mum if there were complete miscibility (reference 11). If the two are very slightly soluble in each other, the activity is practically that of the pure component. If the activity coefficient of the solute is practically that



small electric moment. We may now consider the effect of dipolar ionic structure upon the solubility of the amino acids.

### Melting Points and Crystal Structure of Amino Acids: Their Significance for Solubility

Equation 1 has important implications concerning the ideal solubility of the amino acids. On account of the electrostatic forces binding the molecules in the crystal, their melting points are very high (Table 2). In fact, since they decompose on melting, the melting process is irreversible, and the values experimentally determined represent decomposition points rather than melting points. The true melting points, if they could be determined, would probably be higher than the values reported in Table 2; certainly they are not lower. Thus, since  $T_f$  for most amino acids is near 300°C., or higher, the factor  $\frac{T_f - T}{T}$  in equation 1 is not far from unity for solubility measurements at 25°C. Unless the entropy of fusion for amino acids is abnormally low, this makes their ideal solubility very small. Table 2 gives also the melting points of a number of salts and organic compounds. The values for amino acids are much higher than for most organic compounds, but much lower than for salts, as would be expected from their structure. Among organic compounds which are not dipolar ions, the amides and other substances containing the CONH group have particularly high melting points. Oxamide ( $\text{CO}\cdot\text{NH}_2$ )<sub>2</sub> melts at 419°; glycolamide, the uncharged isomer of glycine, melts at the much lower temperature of 119°.

The significance of the high melting points of amino acids and betaines was clearly recognized by Pfeiffer<sup>12</sup>, who based his conclusion that these compounds were dipolar ions mainly on this line of evidence. Bjerrum<sup>13</sup> also, independently of Pfeiffer, clearly recognized the importance of the evidence on this point in his classical paper on dipolar ions.

Qualitatively the significance of the high melting points of the amino acids is clear. The exact value of the melting point of any substance, however, is a very complex function of the nature of the molecules and their arrangement in the crystal. A deeper understanding of the nature of the forces operative in the crystals can come only by way of determination of the crystal structure through x-ray diffraction measurements. The first investigations in this field were by Bernal<sup>14</sup>, who determined the unit cells and space groups of fifteen amino acids and related compounds. His data were not sufficient to establish the atomic positions in any of these crystals, but he concluded that "for the simpler amino acids the tendency will be to

approach ionic packing (as witnessed by their high specific gravity) . . . the determining factor in the structure of  $\alpha$ -amino acids is the dipole group  $^+\text{H}_3\text{N}\cdot\text{CHR}\cdot\text{COO}^-$ . These groups pack together in double molecules or in extended chains."

In the same years, Hengstenberg and Lenel<sup>15</sup> reported a detailed study of glycine crystals and proposed a structure which, however, has not proved

TABLE 2. MELTING POINTS OF SALTS, AMINO ACIDS AND UNCHARGED MOLECULES

Inorganic Salts	Melting Point (°C.)	Normal Hydrocarbons	Melting Point (°C.)
Barium chloride.....	925	Ethane.....	-172
Sodium chloride.....	804	Propane.....	-189
Potassium chloride.....	772	Pentane.....	-131
Lithium chloride.....	613	Hexane.....	-94
Lithium bromide.....	547	Heptane.....	-90
Lithium iodide.....	446	Octane.....	-56
Amino Acids		Aliphatic Amines and Acids	
Glycine.....	290	Ethylamine.....	-80.6
<i>dl</i> -Alanine.....	297	Propylamine.....	-83.0
<i>dl</i> -Valine.....	292	Amylamine.....	-55.0
<i>dl</i> -Leucine.....	332	Hexylamine.....	-19.0
<i>dl</i> - $\alpha$ -Aminocaproic acid.....	327	Acetic acid.....	16.6
Diglycine.....	262-264	Propionic acid.....	-22.0
<i>dl</i> -Phenylalanine.....	318-320	Valeric acid.....	-34.5
<i>l</i> -Tyrosine.....	342-344	Caproic acid.....	-9.5
Certain Amino Acid Derivatives		Certain other Organic Compounds	
Glycyl amide.....	66-68	Acetamide.....	81
Alanyl amide.....	71-72	Glycolamide.....	119
Leucyl amide.....	105-108	Lactamide.....	77
Formylglycine.....	152-153	Butyramide.....	116
Formylglycylglycine.....	168-170	Benzamide.....	130
Formyl- $\alpha$ -aminobutyric acid.....	154-155	Acetanilide.....	114
Formylleucine.....	114-115	Acetophenone.....	159
Acetylglycine.....	206	Propyl urea.....	154
Hydantoic acid of glycine.....	169-170	Urethane.....	48
Hydantoic acid of alanine.....	169-170	Ethyl allophanate.....	191
Hydantoin of glycine.....	217-218	Glucose.....	146
Hydantoin of $\alpha$ -aminobutyric acid.....	118-120	Sucrose.....	186
Hydantoin of leucine.....	208-209	Urea.....	133
Hydantoin of aspartic acid.....	215-217	Formyl urea.....	169
		Biuret.....	193
		Acetyl urea.....	217
		Carbonyl diurea.....	233
		Allantoin.....	235
		Acetyl biuret.....	194

<sup>15</sup>From Edsall, J. T., in "The Chemistry of the Amino Acids and Proteins," edited by C. L. A. Schmidt, Chapter XVI, p. 894.

to be the correct one. The very thorough and painstaking studies of Albrecht and Corey<sup>16</sup> have finally established the correct structure in detail. The crystal is monoclinic, the unit cell having the dimensions  $a_0 = 5.10 \text{ \AA}$ ,  $b_0 = 11.96 \text{ \AA}$ ,  $c_0 = 5.45 \text{ \AA}$ ,  $\beta = 110^\circ 38'$ ; it contains four molecules<sup>17</sup>. The arrangement is found to consist of "nearly flat glycine

molecules bound together by hydrogen bonds to form discrete layers. . . . Within the molecule interatomic distances are  $C_I - O_I$ , 1.27 Å;  $C_I - O_{II}$ , 1.25 Å;  $C_I - C_{II}$ , 1.52 Å, and  $C_{II} - N$ , 1.39 Å." The probable error in these distances was estimated at  $\pm 0.02$  Å. "The difference, 0.02 Å, between the C - O distances is within the limit of accuracy of their determination, so that no significance is to be attached to it." (Reference 16, p. 1098). The two C and the two O atoms are coplanar, within the limits of error of the measurements. The center of the N atom is 0.268 Å away from the plane of the rest of the molecule. (See Chapter 14, Fig. 3). The authors conclude that "in the glycine crystal the molecule has the 'zwitter-ion' structure,  $H_3N^+CH_2 \cdot COO^-$ , two of the three hydrogen atoms attached to nitrogen forming strong hydrogen bonds to oxygen atoms in the same layer and the third sharing its bond-forming capacity nearly equally between two nearest oxygen atoms in the adjacent layer. The glycine crystal would thus be made up of double layers of molecules held firmly intact by hydrogen bonds between nitrogen and oxygen atoms and also by the electrostatic forces effective between these same atoms which constitute the positive and negative extremes of the polar molecule." (Reference 16, p. 1100).

The influence of dipolar ionic structure on packing within the crystal may be seen by comparing the density of crystalline glycine with that of its uncharged isomer, glycolamide<sup>18</sup>. The density of solid glycine is 1.607 (molal volume 46.71 cc); the density of glycolamide is 1.390 (molal volume 54.01 cc). The difference of 730 cc in the molal volumes reflects the much closer packing in the glycine crystal, because of the electrostatic attraction of the charged groups.

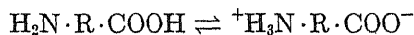
Recently, Levy and Corey<sup>19</sup> have determined in detail the crystal structure of *D*L-alanine. The unit cell is orthorhombic ( $a_0 = 12.04$  Å,  $b_0 = 6.04$  Å,  $c_0 = 5.81$  Å) and contains four molecules of alanine. The dimensions of the molecules are discussed in Chapter 14. The molecules in the crystal are not built into double layers, as in glycine; the additional methyl group prevents the approach of neighboring molecules in the manner which would lead to such a layer structure. Instead, the molecules are linked together by hydrogen bonds and electrostatic forces into a closely knit three-dimensional framework. The positions of the hydrogens are, of course, not precisely fixed by the x-ray analysis; but the structure indicates that the three hydrogens of the  $-NH_3^+$  group are arranged in approximately the tetrahedral positions around the nitrogen, each N—H bond pointing toward an oxygen atom of a  $-COO^-$  group in one of the neighboring molecules. The three atoms (N—H—O) forming each hydro-

gen bond are thus approximately collinear. The two oxygens of the  $\text{—COO}^-$  group in each molecule are different; one is linked by two hydrogen bonds to adjacent nitrogens, the other to only one. The detailed discussion of the structure in the original paper brings out many significant points which cannot be discussed here.

The work of separating an amino acid molecule from the crystal lattice, which is a necessary preliminary to the process of solution, depends primarily on the electrostatic forces between the  $\text{—NH}_3^+$  and  $\text{—COO}^-$  groups and on the hydrogen bonds through which these forces largely operate<sup>20</sup>. The van der Waals forces which are responsible for the cohesive energy of crystals of non-polar molecules are of secondary importance for the amino acids.

#### Solubility and Heat of Solution of Amino Acids in Water

An amino acid in aqueous solution is closely united to some of the immediately surrounding water molecules by forces similar to those which bind it to other amino acid molecules in the crystal. Even if an amino acid had the "uncharged" structure,  $\text{H}_2\text{N}\cdot\text{R}\cdot\text{COOH}$ , the amino and carboxyl groups would form hydrogen bonds with the solvent water. These bonds are very much stronger, however, for a dipolar ion; as is shown, for instance, by the large decreases in volume and heat capacity accompanying the reaction:



These effects, due to the electrostriction of the solvent around the charged groups, have been discussed in Chapter 7. From the values of  $\Delta H$ , and of the equilibrium constant for the above reaction, as given in Chapter 4, we may also estimate that the formation of a dipolar ion in water is accompanied by a large decrease of entropy,  $\Delta S$  being approximately  $-12$  entropy units per mole at  $25^\circ$  for an  $\alpha$ -amino acid, and  $-18$  for an amino acid in which the charged groups are widely separated. This entropy change is a reflection of the increased regularity in the orientation of the water molecules around the amino acid, due to the charged groups of the dipolar ion. All these effects indicate that the cohesive forces between solute and solvent, the magnitude of which is expressed by  $b_{12}$  in equation 4, must be very great. Owing to the magnitude of these forces, it is reasonable that  $\log \gamma$  in that equation should often be small or negative, the solubility of the amino acid being then correspondingly high.

The relative importance of hydrogen bond formation, and of dipolar ionic structure, in determining solubility in water may be seen from

paring three isomers—glycine, glycolamide and nitroethane (Table 3). The first is a dipolar ion; the second is not, although it contains three hydrogens capable of entering into hydrogen bond formation; the third, although highly polar, contains no hydrogen in its polar group. These three isomers show a great decline in melting point, and a very great increase in the alcohol-water solubility ratio, in the order: glycine-glycolamide-nitroethane. The great difference between the two latter compounds, which are about equally polar, confirms the view of Hildebrand (reference 1, Chapter 4) that polarity *alone* cannot form the basis of any adequate theory of solubility.

In recent years, very accurate and concordant solubility values for amino acids have been obtained in at least three different laboratories; by Schmidt and his collaborators (Table 4); by Dunn and his collaborators (Chapter 9, Table 5) and by Cohn, McMeekin and their collaborators (Chapter 9, Tables 2, 3 and 4). The extensive data on solubility in water from Schmidt's laboratory cover a large number of amino acids over a wide temperature range. The individual values are too numerous to be listed

TABLE 3. MELTING POINTS AND SOLUBILITIES OF THE ISOMERS GLYCINE, GLYCOLAMIDE AND NITROETHANE

Substance	Melting Point (°C.)	Sol. in Water, moles per liter at 25°C.	Sol. in Alcohol, moles per liter at 25°C.
$+H_2N \cdot CH_2 \cdot COO^-$ .....	290° (dec)	2.886	0.00039
$HO \cdot CH_2 \cdot CONH_2$ .....	119°	5.509	0.3422
$H_3C \cdot CH_2 \cdot NO_2$ .....	< -50°	0.6 (approx.)	$\infty$

here; the data are compactly summarized by equations which permit the solubility of any amino acid in water to be readily calculated, in most cases for the entire temperature range from 0° to 100° (Table 4).

These data show the enormous variation in solubility with structure, from the very soluble molecules proline, hydroxyproline and glycine to such very insoluble ones as cystine, tyrosine and diiodotyrosine. In the series glycine, alanine, valine, norleucine there is a progressive decrease in solubility with increase in the number of  $-CH_2$  groups<sup>21</sup>; but apart from this regularity, it is difficult to state general relations between solubility and structure. Relative solubility in different solvents, as we shall see, is much more closely related to structure.

The temperature coefficient of solubility is intimately related to the heat of solution of the solute. If we differentiate equation 1 with respect to temperature, we obtain

$$\left( \frac{\partial \ln a_2}{\partial T} \right) = \frac{\Delta H_f/n}{RT^2} \quad (9)$$

The subscript *s* means saturated solution. This differential equation may be applied to the activity relative to any standard state, such as that in which  $\gamma_2$  is unity in a very dilute solution, if  $\Delta H_f/n$  is replaced by the partial molal heat of solution of the solute in that standard state, but the integration constant cannot be determined from the melting point. We may write:

$$\begin{aligned} \left( \frac{\partial \ln a_2}{\partial T} \right)_s &= \left( \frac{\partial \ln N_2}{\partial T} \right)_s + \left( \frac{\partial \ln \gamma_2}{\partial T} \right)_s \\ &= \left( \frac{\partial \ln N_2}{\partial T} \right)_s + \left( \frac{\partial \ln \gamma_2}{\partial \ln N_2} \right)_T \left( \frac{\partial \ln N_2}{\partial T} \right)_s + \left( \frac{\partial \ln \gamma_2}{\partial T} \right)_{N_2} \\ &= \left( \frac{\partial \ln N_2}{\partial T} \right)_s \left[ 1 + \left( \frac{\partial \ln \gamma_2}{\partial \ln N_2} \right)_T \right] + \left( \frac{\partial \ln \gamma_2}{\partial T} \right)_{N_2} \end{aligned} \quad (10)$$

However,  $RT^2 \left( \frac{\partial \ln \gamma_2}{\partial T} \right)_{N_2}$  is the difference between the partial molal heat content of the solute in the standard state and in the saturated solution. Hence

$$RT^2 \left( \frac{\partial \ln N_2}{\partial T} \right)_s \left[ 1 + \left( \frac{\partial \ln \gamma_2}{\partial \ln N_2} \right)_T \right] = \frac{\Delta H_s}{n} \quad (11)$$

where  $\Delta H_s/n$  is the partial molal heat of solution of the solid in a saturated solution<sup>22</sup>. For very insoluble substances we may set  $a_2 = N_2$ , even in the saturated solution.

Table 5 gives data on the heats of solution for amino acids in very dilute and in saturated solutions. These were derived partly from the temperature coefficient of solubility, and partly from calorimetric measurements. The values of  $(\bar{H}_2)_{\text{sat.}} - H_2^0 = RT^2 \left( \frac{\partial \ln \gamma_2}{\partial T} \right)_{N_2}$  for the saturated solution, were derived from measurements of the heat of dilution<sup>23</sup>. The experimental data were all obtained by Schmidt and his collaborators; as tabulated here, the data were recalculated by Borsook and Huffman.

The heat of solution of a solid may be divided into two parts: (1) the heat of sublimation of the crystalline solute, which is determined by its crystal lattice energy, and (2) the heat of solvation of the vaporized solute molecules.

$$\Delta H_{\text{solution}} = \Delta H_{\text{subl.}} + \Delta H_{\text{solv.}} \quad (12)$$

For a liquid solute, the heat of vaporization of course replaces that of sublimation in this equation.

TABLE 4. COEFFICIENTS OF SOLUBILITY EQUATIONS OF CERTAIN AMINO ACIDS IN WATER<sup>a</sup>  
From Borsook, H. and Huffman, H. M.\*

Amino Acid	$a_1$	$b_1 \times 10^2$	$c_1 \times 10^5$	$a_2$	$a_3$	$b_2 \times 10^2$	$c_3 \times 10^5$	$a_4$	$b_4 \times 10^2$	$c_4 \times 10^5$
alanine.....	2.1048	0.4669		0.1551	-2.5792	1.075		-6.5150	1.037	
arginine.....	2.0830	0.5608		0.1333	-3.2199	1.291		-7.1317	1.245	
asparagine·H <sub>2</sub> O.....	0.9289	2.311	-4.981	-1.2475	-25.9584	1.159	-11.47	-30.2463	11.79	-11.84
aspartic acid.....	0.3194	1.519		-1.8047	-23.7113	3.499		-17.7370	3.502	
glutamic acid.....	0.4181	2.016	-4.999	-1.7060	-25.1918	10.93	-11.51	-29.2797	10.98	-11.61
histidine.....	-1.299	1.357		-3.680	-18.643	3.125		-21.023	3.125	
isoleucine.....	-1.7959	0.8013	27.89	-4.1766	19.4912	-23.66	47.61	15.4747	-23.66	47.61
lysine.....	-2.1087	3.367	-22.56	-4.4894	-36.9668	14.48	-16.80	-40.9733	14.48	-16.80
L-cystine <sup>b</sup> .....	-1.7190	0.4514	27.39	-4.0997	34.7268	-33.38	63.01	30.7103	-33.38	63.01
L-cystine <sup>b</sup> .....	-2.6034	5.890	-49.41	-4.9841	-133.4125	75.73	-113.8	-137.429	-75.73	-113.8
proline.....	0.0839	1.627		-2.445	-15.381	3.753		-19.894	3.752	
serine.....	0.188	0.9884		-2.343	-11.610	2.276		-15.537	2.247	
threonine.....	0.0065	1.038	4.648	-2.392	-4.058	-3.450	1.069	-8.426	-3.215	1.030
tryptophan.....	-0.690	1.92		-3.326	-19.745	4.42		-23.761	4.43	
tyrosine.....	-0.827	1.43 <sup>11</sup>		-3.464	-16.989	3.30		-21.006	3.30	
valine.....	0.5331	1.613		-1.6345	-13.9054	3.714		-17.9095	3.709	
isoleucine.....	0.9317	1.523		-1.2359	-12.4244	3.507		-16.4071	3.495	
leucine.....	2.1516	1.087	-4.114	0.2762	-13.2619	7.676	-9.473	-17.8976	8.171	-10.50
methionine.....	2.4603	0.3891		0.3428	-1.6575	0.8959		-5.5906	0.8514	
phenylalanine.....	1.5787	0.07682	2.594	-0.5389	2.7190	-3.081	5.972	-1.3913	-3.020	5.866
proline.....	1.2616	0.2512	3.794	-0.8560	2.9651	-4.193	8.736	-1.1373	-4.134	8.632
serine.....	1.3561	0.02233	3.727	-0.7615	4.5073	-4.683	8.582	-0.7252	-3.814	7.198
threonine.....	0.9013	0.2635	4.591	-1.2163	3.4260	-5.167	10.57	-0.6258	-5.143	10.53
tryptophan.....	1.2597	1.108	1.221	-0.9140	-11.1682	4.086	-2.811	-15.2099	4.111	-2.871
tyrosine.....	0.9258	0.4524	3.402	-1.1918	0.2523	-3.236	7.833	-4.2067	-2.941	7.340
valine.....	1.2974	0.6982		-0.9204	-6.510	1.608		-10.5103	1.601	
phenylalanine.....	0.9986	0.5252	3.140	-1.2192	-0.7184	-2.739	7.229	-5.0876	-2.495	6.808
proline.....	3.1050	0.4206		-0.0441	-0.2407	0.9686		-3.8586	0.7586	
serine.....	1.3432	1.520	-3.548	-0.6782	-17.2153	7.963	-8.169	-21.4529	8.134	-8.504
threonine.....	1.5945	1.916	-8.500	-0.5029	-27.8015	15.10	-19.57	-32.1283	15.35	-20.07
tryptophan.....	0.9156	0.4834	2.988	-1.3942	-11.3824	4.872	6.881	-15.3928	4.869	-6.879
tyrosine.....	-0.708	1.46		-2.966	-10.799	3.36		-20.062	3.37	
valine.....	-0.708	1.46		-2.966	-10.799	3.36		-20.062	3.37	
tryptophan.....	-0.833	1.51		-3.091	-16.562	3.46		-20.577	3.46	

e <sup>d</sup> .....	1.9211	8.1515	8.589	-0.1456	0.6274	-8.927	1.978	-3.4570	-8.528	1.9058
e <sup>d</sup> .....	1.6675	97.75	80.22	-0.4011	-20.8468	123.4	18.47	-24.2293	119.4	17.85
e <sup>d</sup> .....	1.8847	11.42	4.799	-0.1839	-0.3175	-3.406	1.105	-3.6570	-8.125	1.910
e <sup>d</sup> .....	1.9227			-0.1459	-0.3359			-4.3653		
e <sup>d</sup> .....	1.8836			-0.1850	-0.4260			-4.4542		
ae.....	1.7749	0.2389	2.607	-0.2966	-2.2921	-2.729	6.003	-1.7417	-2.705	5.928

ability equations:

- (a)  $\text{Log } S = a_1 + b_1d + c_1t^2$  (grams per 1000 gms. water)  
 (b)  $\text{Log } m = a_2 + b_1t + c_1t^2$  (moles per 1000 gms. water)  
 (c)  $\ln m = a_3 + b_3T + c_3T^2$   
 (d)  $\ln N_2 = a_4 + b_4T + c_4T^2$

e first set of values refer to 273.1° to 303.1° and the second set to 298.1° to 323.5° absolute. (Unpublished data of Campbell, W. W., midt, C. L. A.)

ues are not strictly accurate due to contamination of the leucine with a small amount of methionine.  
 e five sets of values which are given under d-valine refer to the various crystal forms.

nces: Dalton, J. B., and Schmidt, C. L. A., *J. Biol. Chem.*, 103, 549 (1933); *J. Gen. Physiol.*, 19, 767 (1936); Tomiyama, T., and Schmidt, C. L. A., *J. Gen. Physiol.*, 19, 767 (1936); Winnek, P. S., and Schmidt, C. L. A., *J. Gen. Physiol.*, 18, 889 (1933); 19, 773 (1936); see also, Dunn, M. S., Ross, F. J., and Read, L. S., *J. Biol. Chem.*, 103, 579 (1933).  
 8, "The Chemistry of the Amino Acids and Proteins", edited by C. L. A. Schmidt, C. C. Thomas, Springfield, Ill., 1938.



Amino acids may be sublimed with little decomposition at temperatures around 150°<sup>24</sup>, but as yet nothing is known concerning their heats of sublimation. These must certainly be very large, in view of the very powerful forces binding them in the crystal lattice. The process of solvation must

TABLE 5. CALCULATED AND OBSERVED VALUES OF HEATS OF SOLUTION OF AMINO ACIDS IN AQUEOUS SOLUTION AT 25°

Amino Acid	$\bar{H}_2 - H_f^0$ (solid)		$\bar{H}_2$ (sat'd) - $H_f^0$ Measured	$\bar{H}_2$ (sat'd) - $H_2$ (solid)		
	Measured directly	Calculated from solubility data		From thermal data	Calculated from solubility and activity coefficients	$\frac{d \ln \gamma}{d \ln m}$
<i>d</i> -Alanine .....		1830			1830	0
<i>dl</i> -Alanine .....	2040 ±20	2200	210	2250	2200	0
<i>d</i> -Arginine .....	1500		-425 (1.0M)	1100		
<i>l</i> -Asparagine·H <sub>2</sub> O .....	8000	8430		8000		
<i>l</i> -Asparagine .....	5750			5750		
<i>l</i> -Aspartic acid .....	6000	6180	-200?	5800	5580	-0.549
<i>dl</i> -Aspartic acid .....	7100	7160	-200?	6900	6500	-0.549
<i>l</i> -Cystine .....		5510				
<i>l</i> -Dihydroxyrosine .....		7820			7830	0
<i>d</i> -Glutamic acid .....		6540		6300	6050	-0.539
<i>dl</i> -Glutamic acid .....	6530	6170	-200?	6330	5710	-0.539
Glycine .....	3750 ±15	3370	-400	3350	3370	-0.0607
<i>l</i> -Histidine .....	3300		-90 (0.5M)	3200		
<i>l</i> -Hydroxyproline .....	1400	1440	-35 (2M)	1400		
<i>d</i> -Isoleucine .....		840			1780	0
<i>dl</i> -Isoleucine .....		1780			830	0.382
<i>l</i> -Leucine .....		840			2070	0.382
<i>dl</i> -Leucine .....		2000				
<i>d</i> -Lysine .....	-4000 ±100		500 (1M)	-3500		
<i>dl</i> -Methionine .....	4000 ±100	4230		4000		
<i>dl</i> -Norleucine .....		2530			2530	0
<i>l</i> -Phenylalanine .....		2820				
<i>dl</i> -Phenylalanine .....		2760			2760	0
<i>l</i> -Proline .....	-750 ±50	1340	1050 (8M)	>300		
<i>d</i> -Pyroglutamic acid .....	3600 ±40		-1100 (7M)	2500		
<i>dl</i> -Serine .....	5180 ±60	5410	-130	5050		
Taurine .....	6000 ±100	5980	-300	5700		
<i>l</i> -Tryptophane .....		1360				
<i>l</i> -Tyrosine .....		5950			5950	0
<i>d</i> -Valine .....		500				
<i>dl</i> -Valine .....	1430 ±25	1460	300	1730	1590	-0.0549

The data were taken from DeLong, J. B., and Schmidt, C. I. A., *J. Biol. Chem.*, **103**, 549 (1933); 169, 241 (1935); *J. Gen. Physiol.*, **19**, 767 (1936); Farnham, L., and Schmidt, C. I. A., *J. Gen. Physiol.*, **19**, 57 (1936); Zittle, C. A., and Schmidt, C. I. A., *J. Biol. Chem.*, **109**, 241 (1935), and references therein as indicated in the text.

Table and Footnote are taken from Borsook, Henry, and Huffman, Hugh M., in Schmidt, C. I. A., "The Chemistry of the Amino Acids and Proteins," p. 841, Charles C. Thomas, Springfield, Illinois (1938).

involve a large evolution of heat, due to the interaction of the ionic groups of the dipolar ion with the solvent, but the magnitude of this effect must

heat is absorbed on solution, for all the amino acids studied except glycine and proline, it is plain that the heat absorbed on sublimation is generally larger than that evolved in the solvation process. Here again, one term depends only on the properties of the pure solute; the other on the interaction between solvent and solute.

For a solid, such as an amino acid, the heat of sublimation determines the change of the vapor pressure with temperature:

$$\frac{\partial \ln p^0}{\partial T} = \frac{\Delta H_{\text{subl.}}}{RT^2} \quad (13)$$

In general, substances such as the amino acids, which melt at very high temperatures will show very low values of  $p^0$  and high values of  $\Delta H_{\text{subl.}}$  as compared with most organic compounds.

TABLE 6. MOLECULAR COHESION OF VARIOUS ALIPHATIC GROUPS

Group	Molecular Cohesion in cal. per mole	Group	Molecular Cohesion in cal. per mole
$-\text{CH}_3$	1780	$-\text{NH}_2$	3530
$=\text{CH}_2$		$-\text{Cl}$	3400
$-\text{CH}_2-$		less exactly known	
$=\text{CH}-$	990		
$-\text{CH}-$	-380		
$-\text{O}-$	1630	$-\text{F}$	2060
$-\text{OH}$	7250	$-\text{Br}$	4300
$=\text{CO}$	4270	$-\text{J}$	5040
$-\text{CHO}$	4700	$-\text{NO}_2$	7200
$-\text{COOH}$	8970	$-\text{SH}$	4250
$-\text{COOCH}_3$	5600	$-\text{CONH}_2$	13200
$-\text{COOC}_2\text{H}_5$	6230	$-\text{CONH}-$	10600

From Meyer, K. H., and Mark, H., "Der Aufbau der hochpolymeren organischen Naturstoffe", p. 27, Akad. Verl., Leipzig, 1930.

Molal heats of vaporization for many substances may be approximately calculated empirically as a sum of terms assigned to various groups in the molecule. Meyer and Mark<sup>25</sup> have given a table of such cohesive energies for various groups, which is reproduced here (Table 6).

The high values for the hydroxyl and carboxyl groups are again notable; but the very great energy required to vaporize compounds containing  $\text{CONH}_2$  and  $\text{CONH}$  groups is outstanding. The cohesive energy due to the peptide linkage, and to the charged groups of dipolar ions, must play a major role in determining the crystal lattice energy and the solubility of peptides and proteins. Thus Emil Fischer<sup>26</sup> pointed out that the solubility of the glycine peptides in water decreases with increasing number of glycylic residues in the chain. The high symmetry of these peptides and the absence of side chains, permits very close packing of the molecules in the crystal; the high cohesive forces due to the  $-\text{CONH}-$

groups, combined with the electrostatic forces due to the charged groups, make the crystal lattice energy high and the solubility low.

Comparison of the heats of solution of closely related substances may be used to gain some information regarding differences in crystal lattice energy. This is particularly true when optically active and racemic

TABLE 7. HEATS OF SOLUTION AND CRYSTAL DENSITIES OF RACEMIC AND OPTICALLY ACTIVE AMINO ACIDS

Substance		Crystal Density	Heat of Solution $\Delta H$ : cal/mole	$\Delta H_{dl} - \Delta H_{act}$ .
<i>d</i> -Alanine.....	(1)	1.401	1830	
<i>dl</i> -Alanine.....	(1)	1.425	2200	370
<i>d</i> -Valine.....	(2)	1.230	500	
<i>dl</i> -Valine.....	(1)	1.316	1590	1090
<i>l</i> -Leucine.....	(1)	1.165	830	
<i>dl</i> -Leucine.....	(1)	1.191	2070	1240
<i>d</i> -Isoleucine.....	(3)		843	
<i>dl</i> -Isoleucine.....	(1)		1780	937
<i>l</i> -Phenylalanine.....	(3)		2820	
<i>dl</i> -Phenylalanine.....	(1)		2760	-60
<i>l</i> -Tyrosine.....	(4)	1.456	5950	
<i>d</i> -Tyrosine.....	(4)		5950	
<i>dl</i> -Tyrosine.....	(4)		6110	160
<i>dl</i> -Diiodotyrosine.....	(4)		5830	
<i>l</i> -Diiodotyrosine.....	(4)		7830	-2000
<i>l</i> -Aspartic acid.....	(1)		5580	
<i>dl</i> -Aspartic acid.....	(1)		6500	920
<i>d</i> -Glutamic acid.....	(1)	1.538	6050	
<i>dl</i> -Glutamic acid.....	(1)	1.460	5710	-340
$\alpha$ -Amino- <i>n</i> -butyric acid.....		1.231		
$\alpha$ -Amino-isobutyric acid.....		1.278		

The values given here are apparently the most direct values for the different substances.

- (1) Dalton, J. B., and Schmidt, C. L. A., *J. Biol. Chem.*, **103**, 549 (1933).
- (2) Dalton, J. B., and Schmidt, C. L. A., *J. Gen. Physiol.*, **19**, 767 (1935).
- (3) Dalton, J. B., and Schmidt, C. L. A., *J. Biol. Chem.*, **109**, 241 (1935).
- (4) Dalton, J. B., and Schmidt, C. L. A., *J. Gen. Physiol.*, **18**, 889 (1935).
- (5) Edsall, J. T., and Weare, J. H., *J. Am. Chem. Soc.*, **56**, 2270 (1934).
- (6) Zittle, C. A. and Schmidt, C. L. A., *J. Biol. Chem.*, **108**, 161 (1935).

forms of the same compound are considered. The heat of solvation ( $\Delta H_{\text{sol.}}$  in equation 12) is presumably the same for the *d*- and *l*-forms of any substance (provided the solvent is not optically active). There-

Hence the difference in heat of solution between the two forms should reveal directly the difference in their heats of sublimation or crystal lattice energies:

$$(\Delta H_s)_{dl} - (\Delta H_s)_{act.} = (\Delta H_{subl.})_{dl} - (\Delta H_{subl.})_{act.} \quad (14)$$

The available data, summarized in Table 7, reveal that there is generally a marked difference in heat of solution between the *d*- or *l*- and the *dl*-form of an amino acid. In six of nine substances listed, the *dl*-form shows a greater crystal lattice energy; in one (phenylalanine) the two forms are nearly equal; and in two (glutamic acid and diiodotyrosine) the *dl*-form has the lower crystal lattice energy. In most cases the form of lower crystal lattice energy is the more soluble of the two at room temperature; but its solubility increases with temperature less rapidly than that of its isomer. For instance, *d*-alanine is more soluble than *dl*-alanine at 0°, but distinctly less soluble at 100°.

Density also is closely related to crystal lattice energy since high density reflects close packing of molecules in the crystal<sup>27</sup>. Thus, for leucine and valine (Table 7) the racemic form is denser and shows a larger heat of solution than the optically active form. In the case of glutamic acid, the *d*-form is denser than the *dl*-form and shows also higher heat of solution. For the amino acids thus far studied the greater crystal density is always shown by the form possessing the greater heat of solution.

Among other types of isomers,  $\alpha$ -aminoisobutyric acid is markedly denser and also less soluble in water than  $\alpha$ -amino-*n*-butyric acid (Table 7). For such isomers as this,  $\Delta H_{subl.}$  can hardly be considered identical for the two forms, but it is probably nearly the same, and the heat of solution of the iso-acid should herefore be larger than that of its *n*-isomer.

Dalton and Schmidt<sup>28</sup> conclude that all the racemic substances listed in Table 7 are racemic compounds, not mixtures of *d*- and *l*-crystals, with the possible exception of glutamic acid which they consider doubtful. The marked difference in density between the *d*- and *dl*-forms of glutamic acid suggests that it too exists as a racemic compound rather than as a mixture. Loring and du Vigneaud<sup>29</sup>, by studying solubility in the presence of excess *d*- or *l*-cystine, have shown that *dl*-cystine is a racemic compound, and the same method should be applicable to other amino acids.

<sup>27</sup> Cohn, E. J., McMeekin, T. L., Edsall, J. T., and Weare, J. H., *J. Am. Chem. Soc.*, **56**, 2270 (1934).

<sup>28</sup> Dalton, J. B., and Schmidt, C. L. A., *J. Biol. Chem.*, **103**, 549 (1933); **109**, 241 (1935).

<sup>29</sup> Loring, H. S., and du Vigneaud, V., *J. Biol. Chem.*, **107**, 287 (1934).

## Chapter 9

### Interactions between Organic Solvents and Dipolar Ions Estimated from Solubility Ratios

BY EDWIN J. COHN AND JOHN T. EDSALL

#### Relative Solubility in Water and in Organic Solvents

The determination of relative solubility in water and in other solvents provides data for determining the free energy of transfer of the solute from one medium to another. Consider first an ideal experiment in which water and the organic solvent under consideration are separated by a semipermeable membrane, impermeable to both solvents but permeable to the solute, which is present at infinite dilution. At equilibrium, the activity of the solute,  $a$ , in both phases must be equal. Denoting its mole fraction in the aqueous phase by  $N_w^0$ , and in the organic solvent by  $N_A^0$ , and the corresponding activity coefficients by  $f_w$  and  $f_A$  respectively, we have:

$$a = f_w N_w^0 = f_A N_A^0 \quad (1)$$

Since an infinitely dilute solution of the solute in water is taken as the standard state,  $f_w = 1$  by definition, and

$$f_A = N_w^0 / N_A^0 \quad (2)$$

Consider now a saturated solution of the solute in water (at mole fraction  $N_w$ ), and in the organic solvent (at mole fraction  $N_A$ ). *If the same solid phase is in equilibrium with both saturated solutions*, then the activity of the solute in both saturated solutions,  $a_s$ , must be the same as its activity in the solid phase<sup>1</sup>.

$$a_s = N_w f_{w(\text{sat.})} = N_A f_{A(\text{sat.})} \quad (3)$$

Amino acids are relatively insoluble in all organic solvents—so insoluble that we may generally assume that the activity coefficient is independent of the mole fraction of amino acid in such solvents; that is, that  $f_A$  in equation 2 equals  $f_{A(\text{sat.})}$  in equation 3. Thus the activity coefficient of an amino acid in any organic solvent is given by its solubility ratio in water and that solvent, multiplied by its activity coefficient in the saturated

The free energy of transfer of the amino acid from dilute aqueous solution (at mole fraction  $N_w$ ) to the solvent A (at mole fraction  $N_A$ ) is given by the relation:

$$\bar{F}_A - \bar{F}_0 = RT \ln f_A + RT \ln \frac{N_A}{N_w} \quad (5)$$

The values of  $f_{w(\text{sat.})}$  in water are generally not widely different from unity. In the following discussion, therefore, we shall frequently speak of the solubility ratios as if they gave directly the activity coefficient of the amino acid in the organic solvent, although strictly this is true only as a first approximation.

The magnitude of the error involved in this approximation may be seen by examining the values for  $f_{w(\text{sat.})}$  obtained by Smith and Smith<sup>2</sup> from isopiestic vapor pressure measurements at 25°. Their measurements were carried in most cases to approximately saturated aqueous solutions, and they calculated the activity coefficient as  $\gamma = a/m$ , where  $m$  is the molality, (moles per kilogram of solvent). This is not identical with the coefficient,  $f$ , which gives the ratio of activity to mole fraction. If we define both  $\gamma$  and  $f$  as equal to unity at infinite dilution, then at any finite molality the two are related by the equation

$$f = \gamma \left( 1 + \frac{M_s m}{1000} \right) \quad (6)$$

where  $M_s$  is the molecular weight of the solvent. For aqueous solutions this becomes

$$f = \gamma(1 + 0.01802m) \quad (6a)$$

We have calculated values of  $f$  and  $\log f$  from the  $\gamma$  values of Smith and Smith, in the most concentrated solution of each amino acid or peptide reported by them<sup>3</sup>. These are generally approximately saturated solutions, except for the data enclosed in brackets in Table 1. For amino acids whose saturated solutions are less than 3 molal the value of  $\log f_{(\text{sat.})}$  never exceeds  $\pm 0.1$ ; even for glycine  $\log f_{(\text{sat.})} = -0.1122$ ; and these values are very small compared to the logarithms of the solubility ratios of amino acids in water and in organic solvents. Molecules such as proline, sarcosine and betaine, and the extremely soluble amino acids of large dipole moment, like  $\epsilon$ -aminocaproic acid, show much larger values of  $\log f_{(\text{sat.})}$ , and these values are invariably positive. The very high value for betaine is particularly striking, especially when it is remembered that a saturated solution of betaine at 25° is approximately 14 molal, or nearly three times as

concentrated as the most concentrated solution investigated by Smith and Smith. It is obvious that uncorrected solubility ratios for betaine in water and organic solvents would be very far from giving correct activity coefficients.

We may now turn to the solubility studies themselves. Carried out in different laboratories with different preparations and different techniques

TABLE 1. ACTIVITY COEFFICIENTS OF AMINO ACIDS AND PEPTIDES IN APPROXIMATELY SATURATED AQUEOUS SOLUTIONS AT 25°

Substance	<i>m</i>	$\gamma = \frac{a}{m}$	$f = \frac{a}{N}$	log <i>f</i> <sub>sat.</sub>
Glycine..... (1)	3.3	.729	.772	-.1122
<i>dl</i> -Alanine..... (2)	1.9	1.046	1.082	+.0341
<i>dl</i> - $\alpha$ -Amino <i>n</i> butyric acid..... (2)	2.1	1.195	1.240	+.0935
<i>dl</i> - $\alpha$ -Amino- <i>n</i> -Valeric acid..... (2)	0.65	1.072	1.085	+.0352
$\alpha$ -Aminoisobutyric acid..... (2)	1.5	1.177	1.209	+.0824
<i>dl</i> -Valine..... (2)	0.65	1.101	1.114	+.0468
[ $\beta$ -Alanine..... (3)	7.0	1.445	1.627	+.2114]
<i>dl</i> - $\beta$ -Aminobutyric acid..... (3)	7.0	1.790	2.016	+.3044
<i>dl</i> - $\beta$ -Amino Valeric acid..... (3)	6.5	1.870	2.089	+.3200
$\gamma$ -Aminobutyric acid..... (3)	6.5	1.575	1.759	+.2453
<i>dl</i> - $\gamma$ -Amino Valeric acid..... (3)	7.0	2.014	2.268	+.3556
$\epsilon$ -Aminocaproic acid..... (3)	6.5	1.556	1.738	+.2400
Diglycine..... (5)	1.7	0.685	0.706	-.1512
[Triglycine..... (5)	0.3	0.804	0.809	-.0921]
Glycylalanine..... (5)	1.2	0.845	0.863	-.0638
Alanylglycine..... (5)	1.0	0.855	0.871	-.0602
Alanylalanine..... (5)	1.0	1.036	1.055	+.0232
<i>dl</i> -Serine..... (4)	0.5	.907	.915	-.0385
<i>dl</i> -Threonine..... (4)	2.0	.943	.977	-.0101
<i>l</i> -Hydroxyproline..... (4)	2.3	1.034	1.077	+.0321
<i>dl</i> -Proline..... (4)	7.3	2.002	2.265	+.3551
Sarcosine..... (4)	7.0	1.627	1.832	+.2629
[Betaine..... (4)	5.0	3.933	4.288	+.6322]

Values of  $\gamma$  at the indicated concentrations are from the work of P. K. Smith and E. R. B. Smith, references as follows: (1) *J. Biol. Chem.*, 117, 21, 607 (1937); (3) *ibid.*, 132, 47 (1940); (4) 132, 57 (1940); (5) 135, 273 (1940).

Values of *f* were calculated from  $\gamma$  values by the equation:

$$f = \gamma(1 + .01802M)$$

Most of the substances at the indicated concentrations are from the work of P. K. Smith and E. R. B. Smith, references as follows: (1) *J. Biol. Chem.*, 117, 21, 607 (1937); (3) *ibid.*, 132, 47 (1940); (4) 132, 57 (1940); (5) 135, 273 (1940).

The solubility of *l*-proline is approximately 14.0 molal at 25°; see Table 2.)

the results are on the whole extremely satisfactory, especially as a first approximation. In some cases equilibrium has been approached in different ways. The results from different laboratories may be compared in Table 2, in which the solubility in water and ethanol at 25° is reported

TABLE 2. SOLUBILITY OF AMINO ACIDS IN WATER AND ETHANOL AT 25°

Amino acid	Solubility in water m/l		Solubility in ethanol m/l	Solubility in water log $N_0$	Solubility in ethanol log $N_A$	Solubility ratio log $N_A/N_0$
Glycine.....	2.613	(1)				
Glycine.....	2.617	(2)				
Glycine.....	2.885	(3)				
Glycine.....	2.915*	(4)				
Glycine.....	2.886	(5)	0.00039	(5)	-1.247 -1.486 (3)	-4.638 -3.391
<i>d</i> - $\alpha$ -Alanine...	1.557	(2)				
<i>dl</i> - $\alpha$ -Alanine...	1.679	(3)				
<i>dl</i> - $\alpha$ -Alanine...	1.666*	(4)				
<i>dl</i> - $\alpha$ -Alanine...	1.656	(6)		-1.491		
<i>dl</i> - $\alpha$ -Alanine...	1.660	(5)	0.00076	(5)	-1.491	-4.347 -2.856
<i>dl</i> - $\alpha$ -Amino- <i>n</i> - butyric acid.	1.800	(5)	0.00260	(5)	-1.440	-3.818 -2.378
<i>dl</i> - $\alpha$ -Amino- iso-butyric acid.....	1.330	(5)		-1.583 -1.870		
<i>d</i> -Valine.....	0.706	(5)				
<i>dl</i> -Valine.....	0.572	(3)				
<i>dl</i> -Valine.....	0.599*	(4)				
<i>dl</i> -Valine.....	0.571	(5)	0.00128	(5)	-1.967	-4.125 -2.158
<i>dl</i> - $\alpha$ -Amino- <i>n</i> - caproic acid.	0.0877	(7)				
<i>dl</i> - $\alpha$ -Amino- <i>n</i> - caproic acid.	0.0865*	(3)				
<i>dl</i> - $\alpha$ -Amino- <i>n</i> - caproic acid.	0.0890*	(4)				
<i>dl</i> - $\alpha$ -Amino- <i>n</i> - caproic acid.	0.0866	(5)	0.00104	(9)	-2.801	-4.215 -1.414
<i>l</i> -Leucine.....	0.169	(2)				
<i>l</i> -Leucine.....	0.181*	(3)				
<i>l</i> -Leucine.....	0.185	(8)				
<i>l</i> -Leucine.....	0.185	(5)				
<i>l</i> -Leucine.....	0.171	(5)	0.00128	(5)	-2.503	-4.125 -1.622
<i>dl</i> -Leucine.....	0.0741	(1)				
<i>dl</i> -Leucine.....	0.0750	(1)				
<i>dl</i> -Leucine.....	0.0748*	(3)				
<i>dl</i> -Leucine.....	0.0890*	(4)				
<i>dl</i> -Leucine.....	0.0744	(5)		-2.870 -2.249 (10) -2.515 (3)		
<i>d</i> -Isoleucine...						
<i>dl</i> -Isoleucine...						
<i>l</i> -Aspartic acid.	0.0375	(11)	0.0000116	(11)	-3.168	-6.167 -2.999
<i>dl</i> -Aspartic acid .....				-2.978 (3)		
<i>l</i> -Asparagine..	0.186	(11)	0.000023	(11)	-2.468	-5.870 -3.402
<i>d</i> -Glutamic acid.....	0.0585	(11)	0.0000185	(11)	-2.975	-5.967 -2.992
<i>dl</i> -Glutamic acid.....				-2.599 (3)		
<i>dl</i> -Glutamine..	0.291	(11)	0.0000315	(11)	-2.269	-5.735 -3.466
<i>l</i> -Phenylala- nine.....				-2.492 (10)		

ITE OF SC



TABLE 2—(Concluded)

Amino acid	Solubility in water m/l	Solubility in ethanol m/l	Solubility in water log $N_0$	Solubility in ethanol log $N_A$	Solubility ratio log $N_A/N_0$
<i>d</i> -Tyrosine.....			-4.15 (12)		
<i>l</i> -Tyrosine.....			-4.15 (12)		
<i>dl</i> -Tyrosine.....			-4.46 (12)		
<i>l</i> -Diiodotyrosine.....			-4.58 (12)		
<i>dl</i> -Diiodotyrosine.....			-4.85 (12)		
<i>l</i> -Dibromotyrosine.....			-3.84 (12)		
<i>l</i> -Dichlorotyrosine (anhydrous).....			-3.85 (12)		
<i>l</i> -Tryptophane.....			-3.000 (10)		
<i>l</i> -Cystine.....			-5.105 (10, 13)		
<i>d</i> -Cystine.....			-5.105 (13)		
<i>dl</i> -Cystine.....			-5.434 (13)		
Mesocystine.....			-5.376 (13)		
<i>dl</i> -Methionine.....			-2.393 (10)		
<i>l</i> -Proline.....			-0.693 (14)		
<i>l</i> -Hydroxyproline.....			-1.326 (14)		
<i>dl</i> -Serine.....		0.0000639 (15)	-2.065 (10)	-5.427 (15)	-3.362
<i>dl</i> -Threonine <sup>b</sup> .....	1.598 (15)	0.000453 (15)	-1.507	-4.577	-3.070
Sarcosine <sup>c</sup> .....	4.81 (2)		-0.69 (16)	-1.44 (16)	-0.75
Betaine.....			-1.832 (10)		
Taurine.....					

\* An asterisk denotes that results of other laboratories have been recalculated on the basis of our densities.

<sup>a</sup> This density of Schmidt does not refer to his saturated solution, but to a solution containing 23.97 g. of *l*-leucine.

<sup>b</sup> The sample of threonine studied was supplied by Professor W. C. Rose.

<sup>c</sup> The density of this was 20°C.

(1) W. Würzler, *Z. physiol. Chem.*, **97**, 128 (1916).

(2) Ph. Sch. and L. Angern, *O.*, *ibid.*, **133**, 180 (1924).

(3) Dutton, J. P., and Schmidt, C. L. A., *J. Biol. Chem.*, **103**, 549 (1933).

(4) Dunn, M. S., and Schmidt, C. L. A., *ibid.*, **103**, 579 (1933).

(5) Cohn, E. J., Weare, J. H., and Lissall, J. T., and Weare, J. H., *J. Am. Chem. Soc.*, **56**, 2270 (1934).

(6) Holleman, A., *Rec. trav. chim. des Pays-Bas*, **13**, 276 (1894).

(7) Kudielka, *Monatsh.*, **29**, 351 (1908).

(8) Sano, K., *Biochem. Z.*, **168**, 14 (1926).

(9) McMeekin, T. L., Cohn, E. J., and Weare, J. H., *J. Am. Chem. Soc.*, **58**, 2173 (1936).

(10) Dutton, J. P., and Schmidt, C. L. A., *J. Biol. Chem.*, **109**, 241 (1935).

(11) W. McMeekin, and J. H. Weare, *J. Am. Chem. Soc.*, **57**, 626 (1935).

(12) W. McMeekin, and C. L. A., *J. Gen. Physiol.*, **18**, 889 (1935).

(13) Loring, H. S., and Vigneaud, V., *J. Biol. Chem.*, **107**, 267 (1934).

(14) T. L. McMeekin, and C. L. A., *J. Gen. Physiol.*, **19**, 379 (1935).

(15) W. McMeekin, unpublished data.

(16) Stolzenberg, H., *Z. physiol. Chem.*, **92**, 445 (1914).

and their derivatives have been made in formamide and methanol, acetone and butanol. These results are collected in Table 3. For the amino acids solubility is lower in formamide than in water, lower still in all the alcohols, and lowest in acetone. For the strongly polar hydantoic acids

TABLE 3. SOLUBILITY OF AMINO ACIDS AND RELATED SUBSTANCES IN VARIOUS SOLVENTS AT 25°

	Water	Form- amide	Methanol	Ethanol	Butanol	Acetone	Heptanol
Solubility: moles per liter							
Glycine.....(1)	2.886	0.0838	0.00426	0.00039	0.0000959	0.0000305	
$\alpha$ -Alanine.....(2, 3)	1.656			0.00076			
$\beta$ -Alanine.....(1)	6.123			0.00189			
$\alpha$ -Aminocaproic acid.(1)	0.0866	0.0173	0.00854	0.00104	0.000336	0.0000793	
$\epsilon$ -Aminocaproic acid.(1)	3.848	"		0.00194			
Diglycine.....(4)	1.512			0.0000222			
Triglycine.....(1)	1.0229			0.00000106			
Glycolamide.....(5)	5.509			0.342			
Lactamide.....(5)	8.779			2.847			
$\alpha$ -Hydroxycapro- amide.....(5)	0.0830			0.306			
Glycylglycine- amide.....(5)	5.820			0.0980			
Hydantoic acid.....(1)	0.329	0.837	0.0797	0.0242	0.00643	0.00248	
$\alpha$ -Alanine hydantoic acid.....(1)	0.193			0.0440			
$\beta$ -Alanine hydantoic acid.....(1)	0.158			0.0170			
$\alpha$ -Aminocaproic hy- dantoic acid.....(1)	0.00690	0.165	0.1123	0.0477	0.01786	0.00463	
$\epsilon$ -Aminocaproic hy- dantoic acid.....(1)	0.00690			0.00756	0.00278		
Diglycine hydantoic acid.....(1)	0.126			0.00115			
Triglycine hydantoic acid.....(1)	0.0446			0.000077			
Hydantoin.....(4)	0.397			0.0324			
Hydantoin of $\alpha$ -Ami- no- <i>n</i> -butyric acid.(4)	0.863			0.988			
Hydantoin of leucine (4)	0.0124			0.100			
Aspartic acid hydan- toin.....(4)	0.0705			0.0141			
Formylglycine.....(1)	1.849		0.710	0.295			0.0347
Formyl- $\alpha$ -amino-bu- tyric acid.....(1)	0.256		0.646	0.355			0.0500
Formylleucine.....(4)	0.185			1.792			
Solubility: log <i>N</i>							
Glycine.....	-1.247	-2.476	-3.762	-4.638	-5.055	-5.648	
$\alpha$ -Alanine.....	-1.491			-4.347			
$\beta$ -Alanine.....	-0.816			-3.955			
$\alpha$ -Aminocaproic acid.....	-2.801	-3.161	-3.458	-4.216	-4.512	-5.233	
$\epsilon$ -Aminocaproic acid.....	-0.975			-3.947			
Diglycine.....	-1.522			-5.889			
Triglycine.....	-2.241			-7.206			



TABLE 4. SOLUBILITY OF CERTAIN AMINO ACIDS IN WATER, ETHANOL AND ETHANOL-WATER MIXTURES AT 25°

Glycine (1)	dL- Alanine (1, 2)	dL-α- Amino-n- butyric acid (1)	α- Iso- butyric acid (1)	d-Valine (1)	dL- Valine (1)	dL-α- Amino- n-caproic acid (NOR- leucine) (1)	L- Leucine (1)	dL- Leucine (1)	L-Aspartic acid (3)	L-Aspara- gine (3)	β-Ala- nine (4)	ε-Amino- caproic acid (4)	Diglycine (3)	Triglycine (4)
.....	1.656	1.800	1.330	0.706	0.571	0.0866	0.171	0.0744	0.0375	0.186	6.123	3.848	1.512	1.0229
anol..	1.543	1.082	0.775	0.409	0.318	0.0516	0.0977	0.0423	0.0149	0.0750	5.180	3.439	0.531	
anol..	0.507	0.570	0.401	0.231	0.167	0.0346	0.0620	0.0284	0.00675	0.0308	3.531	2.852	0.152	
anol..	0.157	0.260	0.177	0.123	0.0860	0.0271	0.0441	0.0186	0.00294	0.0105	1.687	1.909	0.0862	
anol..	0.0278	0.0359	0.0668	0.0467	0.0280	0.0130	0.0204	0.00848	0.00070		0.242	0.485	0.000774	0.000608
anol..	0.00556	0.00794			0.00922	0.00555	0.00770		0.00021		0.0382	0.0713	0.000463	0.0000381
.....	0.00039	0.00076			0.00128	0.00104	0.00128		0.0000116	0.0000223	0.00189	0.00194	0.0000222	0.00000106

Solubility: moles per liter

Solubility: log N														
.....	-1.491	-1.440	-1.583	-1.870	-1.987	-2.801	-2.503	-2.870	-3.108	-2.468	-0.816	-0.975	-1.522	-2.241
anol..	-1.545	-1.728	-1.780	-2.063	-2.175	-2.971	-2.695	-3.059	-3.513	-2.810	-0.876	-0.998	-1.948	
anol..	-1.910	-1.987	-2.009	-2.282	-2.395	-3.083	-3.827	-3.199	-3.793	-3.135	-1.011	-1.050	-2.436	
anol..	-2.339	-2.334	-2.116	-2.442	-2.599	-3.102	-3.889	-3.266	-4.114	-3.514	-1.283	-1.185	-2.977	
anol..	-2.975	-2.863	-2.750	-2.848	-2.971	-3.304	-3.109	-3.491	-4.580		-1.093	-1.719	-3.845	-4.636
anol..	-3.595	-3.441			-3.376	-3.573	-3.455		-5.019		-2.759	-2.487	-4.648	-5.672
.....	-4.638	-4.347	-3.818		-4.125	-4.215	-4.125		-6.167	-5.870	-3.955	-3.947	-5.889	-7.206

John, E. J., McMeekin, T. L., Edsall, J. T., and Weare, J. H., *J. Am. Chem. Soc.*, **56**, 2270 (1934).  
 Colleman, A. F., and Antusch, A. C., *Rec. trav. chim. Pays-Bas*, **13**, 277 (1894).  
 John, E. J., and Weare, J. H., *J. Am. Chem. Soc.*, **57**, 626 (1935).  
 McMeekin, T. L., Cohn, E. J., and Weare, J. H., *ibid.*, **58**, 2173 (1936).

John, E. J., McMeekin, T. L., Edsall, J. T., and Weare, J. H., *J. Am. Chem. Soc.*, **56**, 2270 (1934).  
 Coleman, A. F., and Antusch, A. C. *Rec. trav. chim. Pays-Bas*, **13**, 277 (1934).  
 McMeekin, T. L., Cohn, E. J., and Weare, J. H., *J. Am. Chem. Soc.*, **57**, 626 (1935).  
 McMeekin, T. L., Cohn, E. J., and Weare, J. H., *ibid.*, **58**, 2173 (1936).

in the two solvents is the same, practically within the limits of experimental error (Table 4).

Such solubility ratios should also be ultimately of biological significance. The organization of the living cell certainly involves aqueous phases of high dielectric constant and fatty phases of low dielectric constant. The distri-

TABLE 5. SOLUBILITY OF CERTAIN AMINO ACIDS IN ETHANOL-WATER MIXTURES AT DIFFERENT TEMPERATURES

Glycine		<i>dl</i> -Serine		<i>dl</i> -Valine		<i>dl</i> -Leucine		<i>dl</i> -Norleucine	
Temp. (°C.)	log <i>N</i>	Temp. (°C.)	log <i>N</i>	Temp. (°C.)	log <i>N</i>	Temp. (°C.)	log <i>N</i>	Temp. (°C.)	log <i>N</i>
Solubility in 20.32 Weight per cent Ethanol									
0.02	-1.9747	0.00	-2.987	0.02	-2.4365	0.00	-3.4067	0.00	-3.3675
24.97	-1.6345	25.14	-2.5243	24.85	-2.2403 <sup>c</sup>	24.97	-3.1124	25.69	-3.0101
44.98	-1.4034	45.15	-2.2154	44.91	-2.0535 <sup>d</sup>	45.24	-2.8761	44.97	-2.7570
65.11	-1.2020	65.26	-1.9393	65.07	-1.8928 <sup>d</sup>	65.16	-2.6478	65.17	-2.5017
Solubility in 42.52 Weight per cent Ethanol									
0.02	-2.4789	0.00	-3.4750	0.02	-2.7986	0.00	-3.6615	0.00	-3.5654
24.97	-2.1007	25.14	-2.9747	24.85	-2.4989 <sup>e</sup>	24.97	-3.2306	25.69	-3.0768
44.98	-1.8327	45.04	-2.6440	44.92	-2.2487 <sup>f</sup>	45.24	-2.9318	44.96	-2.7721
65.10	-1.597	65.25	-2.3655	64.94	-2.0353	65.20	-2.6696	65.17	-2.4695
Solubility in 66.94 Weight per cent Ethanol									
0.02	-3.0915 <sup>a</sup>	0.00	-4.0545 <sup>a</sup>	0.02	-3.1580	0.00	-3.7932 <sup>a</sup>	0.00	-3.6364 <sup>a</sup>
24.97	-2.7423	25.10	-3.6144 <sup>a</sup>	24.93	-2.8327 <sup>g</sup>	24.97	-3.3936	24.97	-3.2125
44.97	-2.5143	45.04	-3.2716	44.92	-2.5867 <sup>g</sup>	45.18	-3.1249 <sup>a</sup>	44.96	-2.9245
65.07	-2.3045	65.24	-3.0357 <sup>a</sup>	64.94	-2.3799	65.15	-2.8697	65.17	-2.6615
Solubility in 92.61 Weight per cent Ethanol									
0.01	-4.3468 <sup>b</sup>	0.00	-5.5229	0.01	-4.0114	0.00	-4.4377	0.00	-4.2197
25.09	-4.0246	25.09	-4.6586	25.04	-3.6990	25.09	-4.0926	25.09	-3.8827
45.19	-3.7932	45.18	-4.6383	45.21	-3.4622	45.18	-3.8297	45.18	-3.6234
65.00	-3.5719	65.01	-4.2262	65.15	-3.2328 <sup>b</sup>	65.07	-3.5768 <sup>b</sup>	65.01	-3.3757

The above table has been composed from Dunn, M. S., and Ross, F. J., *J. Biol. Chem.*, 125, 309 (1938).

<sup>a</sup> Weight per cent ethanol = 67.27.

<sup>b</sup> Weight per cent ethanol = 92.54.

<sup>c</sup> Weight per cent ethanol = 20.62.

<sup>d</sup> Weight per cent ethanol = 20.00.

<sup>e</sup> Weight per cent ethanol = 43.36.

<sup>f</sup> Weight per cent ethanol = 42.66.

<sup>g</sup> Weight per cent ethanol = 67.11.

bution coefficients of many types of molecules between different solvents of these two types must play an important part in the chemical dynamics of the cell; and comparison of the relative solubilities of two different molecules in such solvents as water and alcohol will indicate, at least

**Effect of the CH<sub>2</sub> Groups of Solutes on Solubility Ratios and on Surface Tensions**

The solubility in water of the monoamino-monocarboxylic acids, from glycine to leucine and norleucine, decreases with increasing number of CH<sub>2</sub> groups in the hydrocarbon chain. The aminobutyric acids are exceptions to this rule, being more soluble than alanine. Probably this abnormality is due to their crystal structure and low crystal lattice energy<sup>4</sup>.

If solubility *ratios* in water and organic solvents are considered, however, the effect of the CH<sub>2</sub> group is perfectly regular<sup>5</sup>. The influence of each CH<sub>2</sub> group in the transfer from water to ethanol appears to be the same for the  $\alpha$ -amino acids or for their derivatives. This may be demonstrated on the basis of such data as those in Table 6 by subtracting  $\log N_A/N_0$  (where  $N_A$  is the solubility in mole fraction in ethanol) for glycine from that for alanine and for  $\alpha$ -aminobutyric acid. The differences per CH<sub>2</sub> group are nearly the same, not only for these substances, but for other amino acids of the same family that have been adequately studied, as well as for derivatives of the amino acids and for their isomers, such as the  $\alpha$ -hydroxy-amides, glycolamide and lactamide. The increment in  $\log N_A/N_0$  per CH<sub>2</sub> group, is always positive, and is approximately equal to 0.49. In other words, *each CH<sub>2</sub> group in side chains terminating in a methyl group may be thought of as increasing solubility in ethanol relative to that in water three-fold.*

A general statement of the rule, applicable to several organic solvents, is given in the equation

$$\log N_A/N_0 = K_1 n_{\text{CH}_2} + K_2 \quad (7)$$

where  $n$  is the number of CH<sub>2</sub> groups in side chains ending in methyl groups, and  $K_2$  a constant depending on the nature of the polar group.  $K_2$  may generally be taken as the value of  $\log N_A/N_0$  of the first member of the series<sup>6</sup>; that is, as -3.39 for  $\alpha$ -amino acids; -0.80 for  $\alpha$ -hydroxy-amides; -0.63 for hydantoic acids; and -0.33 for formyl amino acids without paraffin side chains.

It is not generally possible to make computations for the influence of CH<sub>2</sub> groups between polar groups. Thus  $\log N_A/N_0$  is nearly the same for asparagine (-3.402) and glutamine (-3.466). Likewise for aspartic and glutamic acids the values of  $\log N_A/N_0$  are virtually identical, being -2.999 and -2.992 respectively (Table 2), despite the additional CH<sub>2</sub> group of the latter substance. These two pairs of compounds differ only by a CH<sub>2</sub> group which is "shielded" by a COOH or CONH<sub>2</sub> group at the

TABLE 6. INFLUENCE OF STRUCTURE ON SOLUBILITY RATIO IN WATER AND ETHANOL AT 25°

Substances i and k	(log $N_A/N_0$ ) <sub>i</sub>	(log $N_A/N_0$ ) <sub>k</sub>	$\frac{\Delta \log N_A/N_0}{\Delta n}$
Influence of CH <sub>2</sub> group			
Glycine - alanine.....	-3.391	-2.856	+0.54
Glycine - $\alpha$ -aminobutyric acid.....	-3.391	-2.375	+0.50
Glycine - $\alpha$ -aminocaproic acid.....	-3.391	-1.414	+0.49
Valine - leucine.....	-2.158	-1.622	+0.54
Hydantoic - $\alpha$ -alanine hydantoic acid.....	-0.630	-0.137	+0.49
Hydantoic - $\alpha$ -aminocaproic hydantoic acid.....	-0.630	+1.352	+0.49
Glycolamide - lactamide.....	-0.799	-0.254	+0.54
Glycolamide - $\alpha$ -hydroxycaproamide.....	-0.799	+1.084	+0.47
Formylglycine - formylaminobutyric acid....	-0.330	+0.651	+0.49
Formylglycine - formylleucine.....	-0.330	+1.556	+0.47
Influence of CH <sub>2</sub> CONH group			
Glycine - diglycine.....	-3.391	-4.367	-0.98
Glycine - triglycine.....	-3.391	-4.965	-0.79
Hydantoic acid - diglycine hydantoic acid..	-0.630	-1.533	-0.90
Hydantoic acid - triglycine hydantoic acid..	-0.630	-2.253	-0.81
Glycolamide - glycolylglycine amide.....	-0.799	-1.517	-0.72
Influence of OH group			
Acetamide - glycolamide.....	-0.120	-0.799	-0.68
Propionamide - lactamide.....	+0.016	-0.254	-0.27
Caproamide - $\alpha$ -hydroxycaproamide.....	+1.726	+1.084	-0.64
Alanine - serine.....	-2.856	-3.362	-0.51
$\alpha$ -Aminobutyric acid - threonine.....	-2.375	-3.070	-0.69
Influence of methionyl sulfur			
$\alpha$ -Aminovaleric acid - methionine.....	1.90*	-2.444	-0.54
Influence of benzene ring			
Alanine - phenylalanine.....	-2.856	-1.453	+1.40
Influence of dipolar ionization			
Glycine - glycolamide.....	-3.391	-0.799	-2.59
Alanine - lactamide.....	-2.856	-0.254	-2.60
Norleucine- $\alpha$ - hydroxycaproamide.....	-1.414	+1.084	-2.50
Diglycine - glycolylglycine amide.....	-4.367	-1.517	-2.85
Glycine - hydantoic acid.....	-3.391	-0.630	-2.76
Alanine - alanine hydantoic acid.....	-2.856	-0.137	-2.72
$\beta$ -Alanine - $\beta$ -alanine hydantoic acid.....	-3.139	-0.463	-2.68
$\alpha$ -Aminocaproic acid - $\alpha$ -aminocaproic hy- dantoic acid.....	-1.414	+1.352	-2.77
Diglycine - diglycine hydantoic acid.....	-4.367	-1.533	-2.83
Triglycine - triglycine hydantoic acid.....	-4.965	-2.253	-2.71

end of the chain, whereby its power of influencing the solubility ratio is apparently annulled<sup>5, 7, 8</sup>. In other cases as in  $\epsilon$ -aminocaproic acid,  $\text{CH}_2$  groups between polar groups appear to exert some effect in enhancing the solubility ratio, but not the full effect exerted by a side chain of the same length<sup>7</sup>.

The solubility of many of these compounds in a series of organic solvents may also be described by equation 1. Amino acids and hydantoic acids have been studied in formamide and methanol, in butanol and acetone, as well as in ethanol and water. With increase in the number of  $\text{CH}_2$  groups in the alcohol  $K_1$ ,  $[\Delta(\log N_A/N_0)/\Delta(n_{\text{CH}_2})]$ , becomes more positive,  $K_2$  more negative (Table 7).  $K_1$  is numerically largest for heptanol, and  $K_2$  for acetone, among the solvents investigated. These results yield the solubility ratios in a series of organic solvents, in all of which amino acids are relatively insoluble. Solubility in any one of these solvents rather than in water might readily have been adopted as the standard state.

The logarithm of the solubility ratio for several homologous series and several solvents is plotted as a function of the number of  $\text{CH}_2$  groups in Fig. 1.

Similar results have been obtained by a study of the distribution coefficients of amino acids between water and butanol<sup>9</sup>.

This effect of hydrocarbon residues on the solubility ratio reflects the same forces that give rise to the orientation of molecules in surface films<sup>10</sup>. Long-chain fatty acids or alcohols are oriented at an interface so that the polar  $-\text{OH}$  or  $-\text{COOH}$  group dips into the water, while the non-polar hydrocarbon chain is repelled from it. The accumulation of non-polar groupings at the interface lowers the surface tension of the water. "A comparison of various insoluble substances has proved that the spreading tendency depends upon the presence of certain active groups or radicals in the organic molecule, these being the groups which tend to increase the solubility of organic substances in water. For example, pentane,  $\text{C}_5\text{H}_{12}$ , is practically insoluble in water, but amyl alcohol,  $\text{C}_5\text{H}_{11}\text{OH}$ , is relatively soluble. Thus the hydroxyl groups in organic molecules exert strong attractive forces on the hydroxyl groups in the water molecules and these manifest themselves by an increase in solubility. Similarly the carboxyl group,  $\text{COOH}$ , tends to make the lower fatty acids much more soluble in water than the corresponding hydrocarbons" (reference 10, p. 161).

In 1891 Traube<sup>11</sup> stated the rule that with molecules of aliphatic com-

<sup>7</sup> McMeekin, T. L., Cohn, E. J., and Weare, J. H., *J. Am. Chem. Soc.*, **55**, 2075 (1933).

<sup>8</sup> Cohn, E. J., *Chem. Rev.*, **19**, 241 (1938). See also McMeekin, T. L., Cohn, E. J., and Blanchard, M. H., *J. Am. Chem. Soc.*, **59**, 2717 (1937); and *Compt. rend. trav. lab. Cambridge*, **22**, 442 (1939), for a more complete discussion.

<sup>9</sup> England, A., and Cohn, E. J., *J. Am. Chem. Soc.*, **55**, 2081 (1933).



TABLE 7. INFLUENCE OF STRUCTURE ON THE SOLUBILITY RATIO IN WATER AND VARIOUS SOLVENTS AT 25°

Substances i and k	Solubility Ratios		$\frac{\Delta \log N_A/N_0}{\Delta n}$
	$(\log N_A/N_0)_i$	$(\log N_A/N_0)_k$	
Influence of CH <sub>2</sub> group in formamide			
Glycine - $\alpha$ -aminocaproic acid.....	-1.229	-0.360	0.217
Hydantoic acid - $\alpha$ -aminocaproic hydantoic acid.....	+0.755	+1.728	0.243
Influence of CH <sub>2</sub> group in methanol			
Glycine - $\alpha$ -aminocaproic acid.....	-2.515	-0.657	0.465
Hydantoic acid - $\alpha$ -aminocaproic hydantoic acid.....	-0.271	+1.568	0.460
Formyl glycine - formyl $\alpha$ -aminobutyric acid.....	-0.101	+0.786	0.444
Influence of CH <sub>2</sub> group in ethanol			
Glycine - $\alpha$ -alanine.....	-3.391	-2.856	0.535
Glycine - $\alpha$ -aminobutyric acid.....	-3.391	-2.375	0.508
Glycine - $\alpha$ -amino- <i>n</i> -caproic acid.....	-3.391	-1.414	0.494
Hydantoic acid - $\alpha$ -alanine hydantoic acid.....	-0.630	-0.139	0.491
Hydantoic acid - $\alpha$ -aminocaproic hydantoic acid.....	-0.630	+1.352	0.496
Glycine - Valine.....	-3.391	-2.158	0.411
Glycine - Leucine.....	-3.391	-1.622	0.422
Influence of CH <sub>2</sub> group in acetone			
Glycine - $\alpha$ -aminocaproic acid.....	-4.401	-2.432	0.492
Hydantoic acid - $\alpha$ -aminocaproic hydantoic acid.....	-1.520	+0.437	0.489
Influence of CH <sub>2</sub> group in butanol			
Glycine - $\alpha$ -aminocaproic acid.....	-3.808	-1.711	0.524
Hydantoic acid - $\alpha$ -aminocaproic hydantoic acid.....	-1.011	+1.118	0.532
Influence of CH <sub>2</sub> group in heptanol			
Formylglycine - formyl $\alpha$ -aminobutyric acid.....	-0.880	+0.173	0.537
Influence of dipolar ionization in formamide			
Glycine - hydantoic acid.....	-1.229	+0.755	-1.98
$\alpha$ -Aminocaproic acid - $\alpha$ -aminocaproic hydantoic acid.....	-0.360	+1.728	-2.09
Influence of dipolar ionization in methanol			
Glycine - hydantoic acid.....	-2.515	-0.271	-2.24
$\alpha$ -Aminocaproic acid - $\alpha$ -aminocaproic hy-			

TABLE 7—Concluded

Substances i and k	Solubility Ratios		$\frac{\Delta \log N_A/N_0}{\Delta n}$
	$(\log N_A/N_0)_i$	$(\log N_A/N_0)_k$	
Influence of dipolar ionization in ethanol			
Glycine - hydantoic acid.....	-3.391	-0.630	-2.76
$\alpha$ -Alanine - $\alpha$ -alanine hydantoic acid.....	-2.856	-0.139	-2.72
$\alpha$ -Aminocaproic acid - $\alpha$ -aminocaproic hy- dantoic acid.....	-1.414	+1.352	-2.77
Influence of dipolar ionization in acetone			
Glycine - hydantoic acid.....	-4.401	-1.520	-2.88
$\alpha$ -Aminocaproic acid - $\alpha$ -aminocaproic hy- dantoic acid.....	-2.432	+0.437	-2.87
Influence of dipolar ionization in butanol			
Glycine - hydantoic acid.....	-3.808	-1.011	-2.80
$\alpha$ -Aminocaproic acid - $\alpha$ -aminocaproic hy- dantoic acid.....	-1.711	+1.118	-2.83

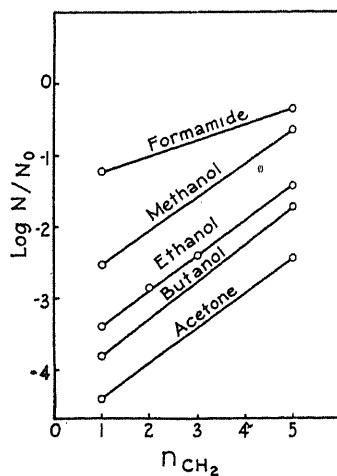


Figure 1

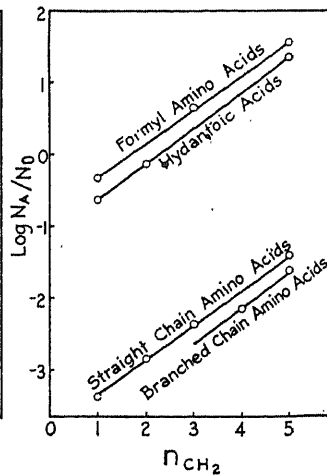


Figure 1a

Figure 1. Solubility ratios of amino acids in various solvents.

Figure 1a. Solubility ratios of amino acids and derivatives in alcohol and water.

pounds having different lengths of hydrocarbon chains, the decrease in the surface tension of the pure liquid, divided by the partial osmotic pressure of the dissolved substance in the underlying solution, for dilute solutions

Lepie and Wyman<sup>12</sup> (see Chapter 7) have shown a similar effect of the CH<sub>2</sub> group on the surface tension of amino acid solutions. The effect of the CH<sub>2</sub> group on vapor pressures and distribution coefficients in the various homologous series shown in Table 7 obviously reflects the same influence.

#### Effect of the CH<sub>2</sub> Groups of Solvents on Solubility Ratios and Activity Coefficients

The results in Table 7 may be further analyzed by dividing the factor  $K_1$  in equation 7 by the concentration of solvent molecules per liter,  $C$ . Values of  $K_1/C$  for various organic solvents are given in Table 8. The increments for each additional CH<sub>2</sub> group become 0.009 for the transfer to formamide; 0.029 to ethanol; 0.036 to acetone and increases to 0.075 to heptanol.

In this series of solvents formamide has the smallest effect upon interactions with both amino acids and hydantoic acids. Formamide has no

Table 8. Influence of CH<sub>2</sub> Groups of Solvent upon CH<sub>2</sub> Groups of Solute Amino Acids or Amino Acid Derivatives

Solvent	CH <sub>2</sub> Groups in side chain of solvent $n$	$K_1$ of solute	$K_1/C$	$\frac{\Delta(K_1/C)^*}{\Delta n}$
Formamide.....	0	0.23	0.0092	
Methanol.....	1	0.46	0.0187	0.0095
Ethanol.....	2	0.49	0.0287	0.0098
Butanol.....	4	0.53	0.0487	0.0099
Heptanol.....	7	0.53	0.0754	0.0095

\* Where  $n$  is the number of CH<sub>2</sub> groups per solvent molecule and  $C$  is the concentration of solvent molecules per liter of solution.

terminal CH<sub>3</sub> group. If we consider it as the first member of the series and divide the increase in the value of  $K_1/C$  by the number of CH<sub>2</sub> groups in the solvent we find a constant increment of slightly less than 0.01 per CH<sub>2</sub> group of the solute for each CH<sub>2</sub> group of the solvent. This relation does not depend upon formamide being considered the first member of the series. Essentially the same result is obtained if the difference in the effect of methanol and ethanol or ethanol and butanol be considered.

These results are computed in this way so that they may be more readily be compared with the change with concentration of activity coefficients derived from freezing point, vapor pressure and solubility measurements, reported in Chapter 10. The solubilities of the amino acids in the organic solvents are not large, and since we are dealing in every case with a change in solubility ratio the high solubilities in water tentatively taken as the

ments reported in Chapter 10, however, the values of  $C$  for the organic solvents are here very large. These solubility measurements could be expected to reflect interactions with organic solvents strictly comparable to those reported in Chapter 10 for dipolar ions only for solubilities that are extremely small and for low values of  $C$ .

#### The Effect of Various Other Groups on Solubility Ratios and Activity Coefficients

The benzene ring, like an aliphatic hydrocarbon radical, may be expected to increase solubility in alcohol relative to that in water. This expectation is confirmed by a comparison of alanine and phenylalanine (Table 6), the relative solubility of the latter in ethanol as compared to water being about 22 times as great as for alanine. On this basis, the benzene ring has approximately the same effect as three  $\text{CH}_2$  groups; but further measurements are needed to determine how generally this relationship holds.

The polar hydroxyl group has the opposite effect. Comparison of alanine and serine, and of  $\alpha$ -amino-*n*-butyric acid with threonine (Table 6) shows that substitution of a hydroxyl group for a hydrogen reduces  $N_A/N_0$  between three and fivefold. Comparison of acetamide and glycolamide, of caproamide and  $\alpha$ -hydroxycaproamide yield results comparable to those deduced from amino acids. The comparisons of hydroxyproline and proline and of tyrosine and phenylalanine are omitted, since they include the effect on non-polar carbon rings of their position between polar groups.

The sulfur of methionine is between non-polar groups, and comparison of this molecule with one of the same composition save for the sulfur suggests that the latter diminishes the influence of the paraffin side chain by an amount approximately equal to one  $\text{CH}_2$  group.

The effect of the  $\text{CONH}$  group on  $\log N_A/N_0$  depends on its position in the molecule. When situated between polar groups, it very markedly diminishes this ratio, as a comparison of the glycine peptides and of their hydantoic acids definitely indicates. The substitution of a  $\text{CONH}_2$  group for a hydrogen atom at the end of a chain, however, appears to produce only a very slight effect on the solubility ratios in compounds possessing other polar or charged groups. The comparison of hydantoic acid ( $\text{HOOC}\cdot\text{CH}_2\cdot\text{NH}\cdot\text{CONH}_2$ ) and glycolamide ( $\text{HO}\cdot\text{CH}_2\cdot\text{CONH}_2$ ) whose net structural difference amounts to one  $\text{CONH}$  group, indicates a change in  $\log N_A/N_0$  of only +0.169 due to this difference; and study of other related compounds containing a terminal  $\text{CONH}_2$  group indicates that its effect on the solubility ratio is generally as small or smaller than this<sup>7</sup>

pattern in the peptide. The influence of this configuration is opposite in sign to that of the  $\text{CH}_2$  group—see Table 6—and is slightly greater than the effect of an hydroxyl group. These estimates of the influence of the various groups of amino acids, peptides and proteins upon  $\Delta \log N_A/N_0$  may be divided (as are the comparable estimates in Table 8) by the moles of solvent per liter,  $C$ . The ratio  $K_1/C$  is given in Table 9, which summarizes these effects.

The same measurements are employed in the last column of the table in calculating the work involved in the transfer of a group of each kind to each solvent investigated. These results suggest that the free energy increment for the  $\text{CH}_2$  group in aqueous solutions referred to alcohols, ranges

Table 9. Influence of the Structure of the Solvent upon the Solubility Ratio in Water and Organic Solvents at 25°

Group of solute	Organic solvent	Dielectric constant	$K_1$ of solute	$K_1/C$	$\Delta F$ calculated*
$\text{CH}_2$	Formamide	84	0.23	0.0092	314
	Methanol	32.71	0.46	0.0187	600
	Ethanol	24.28	0.49	0.0287	668
	Butanol	17.51	0.53	0.0487	723
	Heptanol	9.33	0.53	0.0754	723
	Acetone	20.83	0.49	0.1013	668
Benzene ring	Ethanol	24.28	1.40	0.082	1908
Methionyl sulfur	Ethanol	24.28	-0.54	-0.032	-736
OH	Ethanol	24.28	-0.66	-0.039	-900
$\text{CH}_2\text{CONH}$	Ethanol	24.28	-0.84	-0.047	-1145
Dipolar ionization	Formamide	84	-2.03	-0.0808	-2768
	Methanol	32.71	-2.24	-0.0912	-3054
	Ethanol	24.28	-2.73	-0.161	-3722
	Butanol	17.51	-2.82	-0.259	-3845
	Acetone	20.83	-2.87	-0.593	-3913

\* Change in free energy with change in solvent. See also Cohn, Chem. Rev. 19, 241 (1936) Table 3.

from 600 to 723 calories. Butler<sup>13</sup> estimates the free energy in aqueous solution referred to the gaseous state as 200 calories per  $\text{CH}_2$  group. Butler further estimates a change in free energy of about 6000 calories if  $\text{NH}_2$  is substituted, and 7400 calories if  $\text{COOH}$  is substituted in the paraffin chain.

#### Influence of Dipolar Ionization on Solubility Ratios and Activity Coefficients

Dipolar ions are far more soluble in water, relative to organic solvents

by a comparison of simple amino acids with the isomeric hydroxyamides, with which they have already been compared in evaluating the electrostriction effect (Chapter 7). The amino acids are approximately 400 times as soluble in water, relative to ethanol, as the hydroxyamides. The values of  $\Delta \log N_A/N_0$  from the comparison of glycine-glycolamide and alanine-lactamide are respectively  $-2.59$  and  $-2.60$ .

A similar comparison may be made between amino acids and peptides, on the one hand, and the corresponding hydantoic acids on the other<sup>7</sup>. The comparison for the three hydantoic acids and the  $\alpha$ -amino acids from which they are derived yields  $-2.76$ ,  $-2.72$  and  $-2.77$ , or as an average,  $-2.75$ . Each dipolar ion differs by a terminal CONH group from its hydantoic acid, but it has already been pointed out that the effect of this group on the solubility ratio is apparently small ( $+0.17$  or less in its effect on  $\log N_A/N_0$ ), and furthermore it is presumably essentially the same throughout the whole series.

The influence of dipolar ionization is on the whole remarkably constant in each solvent. The extensive data for ethanol show at most a small influence of the dipole moment of the dipolar ion, even for substances differing as widely in moment as glycine and triglycine. The comparison of diglycine and its isomer glycolylglycineamide, given in the last column of Table 6, in the same terms employed above, yields  $-2.85$ , and that of diglycine and of triglycine and their hydantoic acids  $-2.83$  and  $-2.71$ , respectively. This virtual independence of the dipole moment of the dipolar ions could not occur if the observed differences were determined entirely by electrostatic forces; for in this case the effect of dipolar ionization should be greatest for the substances of greatest dipole moment.

Thus, in the extension of Debye's treatment of ions to the case of dipolar ions if the transfer is from water to another medium of dielectric constant,  $D$ , an equation for the free energy of the transfer would have the form:

$$\bar{F}_e - \bar{F}_e^0 = K_2 \left( \frac{1}{D} - \frac{1}{D_0} \right) \quad (8)$$

For the case of ions, treated by Debye and McAulay<sup>14</sup>,  $K_2$  equals  $(N \epsilon^2 z^2 / 2b)$ , where  $z$  is the valence of the ion and  $b$  its radius. In their first tentative treatment of a dipolar ion as a molecule made up of two spheres of radius  $b$ , whose centers are separated by a distance,  $R$ , Scatchard and Kirkwood<sup>15</sup> deduced that  $K_2$  would be equal to  $N \epsilon^2 z^2 (1/b - 1/R)$ . In Kirkwood's treatment<sup>16</sup> of a dipolar ion as a sphere of radius  $b$ ,  $R$  again being the distance between the charges, the value of  $K_2$  becomes

$$K_2 = \frac{1}{2} N \epsilon^2 z^2 \left( \frac{1}{b} - \frac{1}{R} \right) = \frac{1}{2} N \epsilon^2 z^2 \left( \frac{1}{b} - \frac{1}{D} \right)$$

where  $l$  is the perpendicular distance from the center of the sphere to the line connecting the two charges; that is, the distance from the center of the sphere to the center of the dipole.

In both of these models the change in free energy with change in solvent thus increases markedly with change in  $R$ , whereas in the experiments reported in Tables 6 and 7 no such relation to dipole moment is apparent.

On the assumption that the change in free energy due to electrostatic forces could be estimated from the observed solubility ratios for dipolar ions only after correction for the specific effects of the groups considered in Tables 6 and 7, an equation might be developed incorporating the rule formulated by equation 7:

$$\overline{F}_e - \overline{F}_e^0 = -2.303kT (\log N/N_0 - K_1 n_{\text{CH}_2}) = K_2(1/D - 1/D_0) \quad (9)$$

$K_2$ , however, shows no increase with increase in the dipole moment of the dipolar ion. The electrostatic treatment must therefore be considered far less satisfactory for the interaction of organic solvents with dipolar ions than with ions. This conclusion is confirmed by a consideration of the results in Table 9. Thus formamide, which has a dielectric constant greater than that of water ( $>84$ ), should, on purely electrostatic grounds, give a higher value of  $\log N/N_0$  for dipolar ions than for uncharged compounds. The opposite is in fact the case; the effect of dipolar ionization is the same in direction, although smaller in magnitude, than in the solvents of very low dielectric constant. Evidently the differences between the two classes of compounds are determined in large measure by other forces. Specific interactions between the different chemical groups in both solvents and solutes must clearly be taken into account in addition to the electrostatic forces.

### Solubility in Ethanol-Water Mixtures

The solubility of a group of charged and uncharged organic molecules in alcohol-water mixtures is shown graphically in Figs. 2 and 3. At one extreme in Fig. 2 is acetnaphthalide, with a polar residue attached to the very large non-polar naphthalene ring; at the other is triglycine, with a very high dipole moment and two peptide linkages; and the intervening series of molecules forms a progressive series between the two extremes. All the molecules shown are fairly polar; a truly non-polar molecule like hexane, would doubtless show a curve rising far more steeply even than that of acetnaphthalide.

It is extremely difficult to give a theoretical treatment of the solubility of amino acids and other organic compounds in mixed solvents. The

water molecules tend to cluster closely around the charged groups, displacing the molecules of an organic solvent such as ethanol, which contains a large fraction of non-polar groups in the molecule. Thus the composition

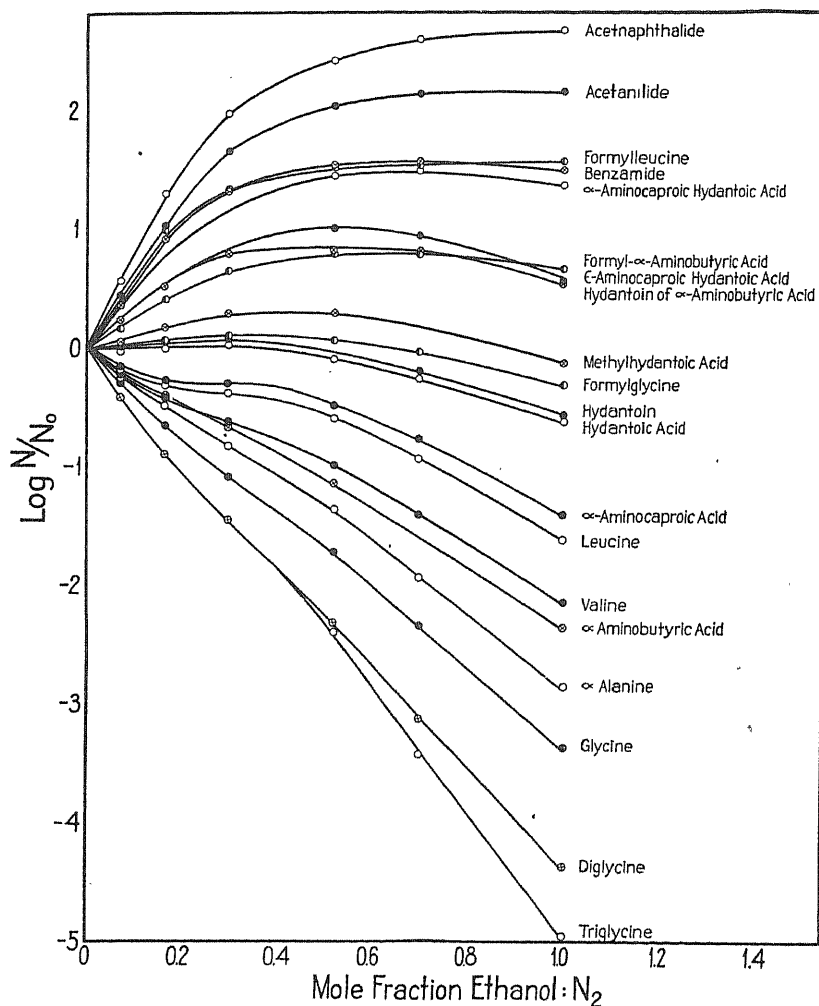


Figure 2. Solubility of amino acids, peptides and related substances in ethanol-water mixtures at 25°

of the solvent varies greatly from point to point in the region surrounding the solute molecules. The free energies and other thermodynamic func-

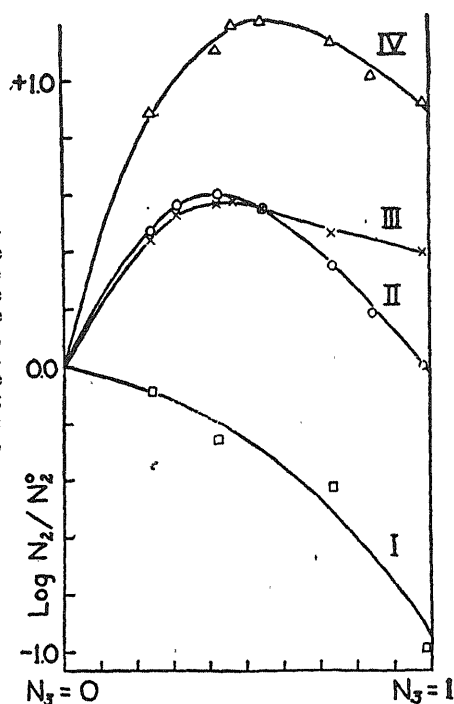


tensive tabulations of the solubilities of amino acids in alcohol-water mixtures given in Tables 4 and 5 supply data of practical utility for chemists concerned with the crystallization of amino acids in such media. In the future they should be amenable to theoretical treatment and we may look forward to such a theoretical analysis.

Notable is a group of compounds which are more soluble in ethanol-water mixtures than in pure water or ethanol; such substances as the dihalogenated tyrosines<sup>17</sup>, methylhydantoic acid,  $\epsilon$ -aminocaproic hydantoic

Figure 3.

The relationship between the logarithm of the solubility ratio and the mole fraction of ethanol in the alcohol-water mixture for *l*-tyrosine (Curve I), *l*-dibromotyrosine (Curve II), *l*-dichlorotyrosine (Curve III), and *l*-diiodotyrosine (Curve IV). Maximum error in data: *l*-tyrosine and *l*-diiodotyrosine  $\pm 5$  per cent; *l*-dibromo- and *l*-dichlorotyrosine  $\pm 1.6$  per cent. From Winnek, P. S., and Schmidt, C. L. A., *J. Gen. Physiol.*, 19, 773 (1936).



acid, and formylglycine. All these molecules contain uncharged polar groups; the dihalogenated tyrosines, which are dipolar ions, contain also the benzene ring with its phenolic groups and two halogen atoms. Apparently molecules of this sort exert attractive forces on both water and alcohol molecules which are more fully satisfied in a mixture of these two solvents than in either one alone. The same type of behavior, in a more extreme form, is found among a class of proteins, the prolamines.

<sup>17</sup> Winnek, P. S., and Schmidt, C. L. A., *J. Gen. Physiol.*, 19, 773 (1936).

## Chapter 10

### Interactions between Amino Acids, Peptides and Related Substances

BY EDWIN J. COHN

An adequate theory of solutions for biological systems must depend upon the development of laws defining the mutual interactions of ions and such dipolar ions as peptides, phospholipids, and proteins. Unquestionably the most important dipolar ions, as their name implies, are the proteins. In order to increase our knowledge of this class of molecules, we have found it desirable, however, first to consider comparable investigations upon smaller molecules of known structure which contain the same reactive groups.

The laws of ideal solution generally hold only for very dilute solutions and deviations from these laws are conveniently expressed in terms of activity coefficients. The activity of a solute may be measured directly, as for instance by EMF measurements on solutions of electrolytes, employing electrodes reversible to the ions concerned; or it may be determined indirectly, by measurement of the activity of the solvent, since the activities of solvent and solute are thermodynamically related. Determinations of activity from freezing point lowering, boiling point elevation and osmotic pressure fall in the latter category. So do direct determinations of the vapor pressure of the solvent, as a function of solute concentration. Although osmotic pressure determinations have proved of particular value in the study of proteins, freezing point and vapor pressure measurements yield more accurate results in systems composed of smaller molecules, such as the amino acids or peptides. Solubility measurements have proved valuable for both large and small molecules.

#### Activity Coefficients of $\alpha$ -Amino Acids as a Function of Concentration

Studies have now been carried out upon the activity coefficients of amino acids as a function of concentration. The influence of glycine in lowering the freezing point of aqueous solutions has been accurately investigated by Szentpály and Prentiss.<sup>1</sup> Their results, calculated as activity coeff

where  $m$  is concentration in moles per 1000 grams of water. The temperature coefficient of  $\log \gamma$  may be calculated from measurements of heat capacities and heats of dilution<sup>2</sup>. From the data of Zittle and Schmidt<sup>2</sup>, Scatchard calculates for glycine at 25°

$$-\log \gamma = 0.08366m - 0.01507m^2 \quad (2)$$

The activity coefficients of glycine have also been determined by vapor pressure measurements by Smith and Smith<sup>5</sup> and by M. M. Richards<sup>6</sup>,

TABLE 1. ACTIVITY COEFFICIENTS OF GLYCINE IN AQUEOUS SOLUTION AT 25°C.

Glycine concentration		Dielectric constant $D$	-Log $\gamma$ (Smith and Smith, <sup>5</sup> by vapor pressure)	-Log $\gamma$ (Richards, <sup>6</sup> by vapor pressure)	-Log $\gamma$ (Scatchard and Prentiss <sup>1</sup> by freezing point)	-Log $\gamma$ (calculated†)
$m$	$C$					
<i>moles per 1000 grams</i>	<i>moles per liter</i>					
0.1	0.099	80.8	0.0088		0.0082	0.0089
0.2	0.198	83.0	0.0168	0.0177	0.0161	0.0173
0.3	0.296	85.2	0.0241		0.0237	0.0252
0.4	0.393	87.4	0.0315	0.0330	0.0310	0.0326
0.5	0.489	89.6	0.0386		0.0381	0.0395
0.6	0.585	91.7		0.0463		0.0460
0.7	0.679	93.8	0.0516		0.0512	0.0521
0.8	0.773	96.0		0.0581		0.0578
1.0	0.958	100.1	0.0685	0.0686†	0.0686	0.0684
1.2	1.140	104.2	0.0768	0.0780	0.0787	0.0780
1.4	1.318	108.3		0.0865		0.0863
1.5	1.406	110.3	0.0883		0.0916	0.0903
1.6	1.494	112.2		0.0941		0.0939
1.7	1.580	114.2	0.0964		0.0987	0.0976
1.8	1.666	116.1		0.1009		0.101
2.0	1.836	120.0	0.105	0.107	0.107	0.107
2.2	2.002	123.7		0.113		0.113
2.4	2.166	127.5		0.119		0.118
2.5	2.247	129.3	0.119			0.121
2.6	2.327	131.1		0.124		0.123
2.8	2.485	134.7		0.128		0.128
3.0	2.640	138.3	0.130	0.133		0.132
3.2	2.793	141.8		0.136		0.135
3.3	2.869	143.5	0.137			0.137

\* Calculated for 25° by means of equation (2).

† Calculated by means of

$\log \gamma = K_R^* (D_0/D) - K^*$ , where  $K_R^* = 0.100$  and  $K^* = 0.007$ .

‡ This value is assumed equal to that of Scatchard and Prentiss at the same temperature and concentration (see reference 6, page 733).

From Cohn, E. J., McMeekin, T. L., Ferry, J. D., and Blanchard, M. H., Ref. 13.

and their results are in good agreement as a first approximation\* with the freezing point determinations of Scatchard and Prentiss<sup>1</sup>, with which they are compared in Table 1.

Several qualitative observations may be made regarding these data:

<sup>2</sup> Zittle, C. A., and Schmidt, C. L. A., *J. Biol. Chem.*, 108, 161 (1935).  
Gucker<sup>3</sup> and Scatchard<sup>4</sup> have determined the heat capacities and the heats of dilution of glycine, alanine, and of certain related molecules, and these results when combined should lead to slight revision of these

(a) as in the case of most electrolytes, the activity coefficients are less than unity; (b) unlike the case of electrolytes,  $-\log \gamma$  is proportional in dilute solution to the concentration,  $C$ , and not to its square root; and (c) the ratio  $-(\log \gamma)/C$  does not remain constant but increases with diminishing concentration, approaching a value close to 0.10 at infinite dilution.

**Influence of  $\text{CH}_2$  Groups.** The electric moments of glycine and alanine are presumably identical. Alanine differs from glycine by substitution of a hydrogen by a methyl group. The volume of the molecule is thus larger by one  $\text{CH}_2$  group or by 16.3 cc. per mole<sup>7</sup>, and the center of the dipole is presumably farther from the center of the molecule than in the case of glycine<sup>8, 9, 10</sup>. Interaction between alanine molecules has now been

TABLE 2. ACTIVITY COEFFICIENTS OF  $\alpha$ -AMINO ACIDS IN AQUEOUS SOLUTION AT 25° FROM THE MEASUREMENTS OF SMITH AND SMITH<sup>5, 11</sup>

Moles per 1000 g. water <i>m</i>	<i>dl</i> -Alanine -log $\gamma$	Sar-cosine* -log $\gamma$	<i>dl</i> -Serine -log $\gamma$	<i>dl</i> - $\alpha$ -amino- <i>n</i> -butyric acid* -log $\gamma$	<i>dl</i> -Threonine -log $\gamma$	$\alpha$ -Amino-iso-butyrac acid -log $\gamma$	<i>dl</i> - $\alpha$ -amino- <i>n</i> -valeric acid -log $\gamma$	<i>dl</i> -Valine -log $\gamma$	Be-taine* -log $\gamma$	Proline* -log $\gamma$	<i>l</i> -Hy-droxy-proline -log $\gamma$
0.0											
0.2	-0.002	-0.002	0.016	-0.009	0.005	-0.011	-0.009	-0.013	-0.030	-0.008	0.000
0.3	-0.003	-0.003	0.025	-0.012	0.007	-0.016	-0.014	-0.019	-0.044	-0.012	0.000
0.4	-0.004			-0.017		-0.021	-0.019	-0.025			
0.5	-0.005	-0.005	0.042	-0.020	0.011	-0.026	-0.023	-0.032	-0.074	-0.020	-0.001
0.65	-0.006			-0.026		-0.033	-0.030	-0.042			
0.7		-0.009			0.015				-0.104	-0.029	-0.001
0.8	-0.008			-0.032		-0.041					
1.0	-0.010	-0.014		-0.040	0.018	-0.050			-0.147	-0.040	-0.003
1.2	-0.012	-0.018		-0.047	0.020	-0.058			-0.176		-0.004
1.5	-0.015	-0.025		-0.057	0.022	-0.071			-0.219	-0.060	-0.006
1.7	-0.017			-0.064							
1.9	-0.020			-0.071							
2.0		-0.038			0.025				-0.289	-0.081	-0.011

\* Activity coefficients are reported for higher concentrations than 2.0 molal in the original communication.

studied by measurements both of vapor pressure<sup>11</sup> and of freezing point<sup>9</sup>. The vapor pressure measurements of Smith and Smith<sup>11</sup>, reported in Table 2, are plotted in Fig. 1.

Whereas the activity coefficients of glycine, like those of electrolytes, are smaller than unity in dilute solution, those of alanine are greater. Moreover,  $-(\log \gamma)/C$  is nearly independent of alanine concentration\*,

<sup>7</sup> Cohn, E. J., McMeekin, T. L., Edsall, J. T., and Blanchard, M. H., *J. Am. Chem. Soc.*, **56**, 784 (1934).

<sup>8</sup> Cohn, E. J., *Chem. Rev.*, **19**, 241 (1936).

<sup>9</sup> Kirkwood, J. G., Personal communication.

<sup>10</sup> Cohn, E. J., *Annual Review of Biochemistry*, **4**, 93 (1935).

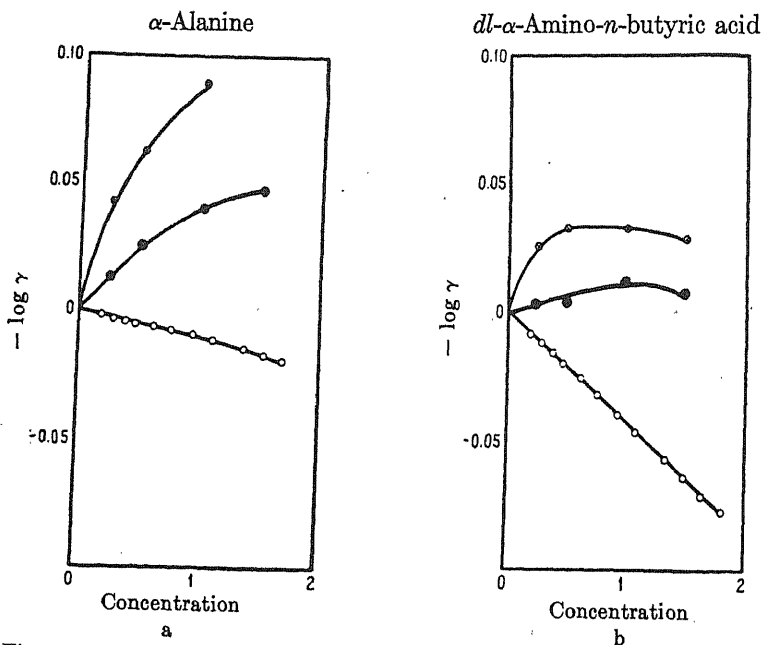
<sup>11</sup> Smith, P. K., and Smith, E. R. B., *J. Biol. Chem.*, **121**, 607 (1937).

\* The vapor pressure measurements of Smith and Smith<sup>5, 11</sup> are reported as osmotic coefficients,  $\phi$ , and empirical equations given for the variation in terms of concentration in moles per 1000 grams of water,  $m$ , of  $\phi$  and  $\gamma$ . The ratio  $-(\log \gamma)/m$ , calculated from their results, diminishes with increase in concentration. The concentration as moles per liter,  $C$ , was estimated from the relation:

$$M = 1000C/(1000 - \phi C)997.08$$

having a value of  $-0.005$ , as compared with a value for glycine (Fig. 1) which (a) is opposite in sign, (b) is nearly ten times as great as for alanine, and (c) diminishes with increase in concentration. If the effect for glycine is ascribed to the dipole moment of its molecules, then that for alanine must be ascribed to large repulsive forces in aqueous solution between the dipole of one molecule and the non-polar side chain of a second.

This effect is far greater if the side chain of an  $\alpha$ -amino acid is an ethyl group, as in  $\alpha$ -aminobutyric acid (Fig. 1), and still greater if the paraffin chain is still longer.



Figures 1a and 1b. Dipolar Ions in Solutions of Alanine and  $\alpha$ -Aminobutyric Acid.

The values denoted by  $\circ$  are from the vapor pressure measurements of Smith and Smith (5). Other values from reference 13.

1a.  $\otimes$ , cystine;  $\bullet$ , asparagine;  $\circ$ , alanine.

1b.  $\otimes$ , cystine;  $\bullet$ , asparagine;  $\circ$ ,  $\alpha$ -aminobutyric acid.

The slope  $-(\log \gamma)/C$  is  $-0.005$  for  $dl$ -alanine,  $-0.033$  for  $dl$ - $\alpha$ -amino- $n$ -butyric acid, and  $-0.047$  for  $dl$ - $\alpha$ -aminovaleric acid. Clearly the effect does not continue to increase at the same rate with further increase in the length of the paraffin side chain. It would appear to be somewhat greater for branched than for straight paraffin chains,\* according to the measure-

\* The change in free energy in the transfer from water to ethanol is also greater for branched than for straight paraffin chains.

ments of Smith and Smith<sup>11</sup>, since the limiting slope is  $-0.054$  for  $\alpha$ -amino-isobutyric acid and  $-0.064$  for *dl*-valine.

These estimated limiting slopes are recorded in Table 3, and the influence of each  $\text{CH}_2$  group in side chains ending in methyl groups estimated by subtracting values of  $-(\log \gamma)/C$  of dipolar ions differing only in the length of their paraffin side chains from the previous member of each series. The increment per  $\text{CH}_2$  group is always negative, but diminishes from approximately  $-0.10$  for the change from glycine to alanine to a value about one-tenth of this, or  $-0.01$  for the change from either the straight or branched chain  $\alpha$ -aminobutyric acid to the comparable  $\alpha$ -aminovaleric acid.

**Influence of OH Groups.** Amino acids differ from each other not only with respect to methyl groups, but also with respect to other configurations. The influence of the hydroxyl group upon the activity coef-

TABLE 3. INFLUENCE OF STRUCTURE ON ACTIVITY COEFFICIENTS OF  $\alpha$ -AMINO ACIDS

Substances <i>i</i> and <i>k</i>	$(-\log \gamma)/C)_i$	$(-\log \gamma)/C)_k$	$\frac{\Delta - (\log \gamma)/C}{\Delta n}$
Influence of $\text{CH}_2$ Group			
Glycine—alanine .....	0.096	-0.005	-0.101
Sarcosine—betaine .....	-0.008	-0.150	-0.071
Serine—threonine .....	0.075	0.027	-0.048
Alanine— $\alpha$ -amino- <i>n</i> -butyric acid .....	-0.005	-0.033	-0.028
$\alpha$ -Aminobutyric— $\alpha$ -aminovaleric acid .....	-0.033	-0.047	-0.014
$\alpha$ -Aminoisobutyric— $\alpha$ -Aminoisovaleric acid .....	-0.054	-0.064	-0.010
Influence of Hydroxyl Group			
<i>dl</i> - $\alpha$ -Alanine— <i>dl</i> -serine .....	-0.005	0.075	0.080
<i>dl</i> - $\alpha$ -Amino- <i>n</i> -butyric acid— <i>dl</i> -threonine .....	-0.033	0.027	0.060
<i>dl</i> -Proline— <i>l</i> -hydroxyproline .....	-0.043	-0.001	0.042

ficients of amino acids has also been investigated by Smith and Smith<sup>11</sup> and their results are given in Table 2. These results for serine and threonine, each of which has a hydroxyl group, may be employed in estimating the influence of the OH group. The polar nature of the hydroxyl group may be seen (Table 3) to result in a limiting slope for serine greater than that for alanine and for threonine greater than that for  $\alpha$ -amino-*n*-butyric acid. The difference in the values of  $-(\log \gamma)/C$  for serine and threonine is also greater than for alanine and  $\alpha$ -amino-*n*-butyric acid, thus indicating that the effect of  $\text{CH}_2$  groups in diminishing the values of  $-(\log \gamma)/C$  is greater the more polar the molecule in which they are substituted.

$\alpha$ -Aminobutyric acid .....	-2.375	-3.355
$\alpha$ -Aminocaproic acid .....	-1.414	-3.374

Whereas methyl groups increase, hydroxyl groups diminish the activity coefficients of the amino acids. The influence of the hydroxyl group is estimated in Table 3 by comparing alanine and serine,  $\alpha$ -amino-*n*-butyric acid and threonine, proline and hydroxyproline. These results contrast interestingly with those in Table 6 of Chapter 9. The negative logarithm of the activity coefficient, (as judged from the logarithm of the solubility ratio,  $\log N/N_0$ ) of an amino acid in an organic solvent is negative in sign and greater, the greater the number of  $\text{CH}_2$  groups in the organic solvent, by 0.01 per  $\text{CH}_2$  group per mole of solvent. Each  $\text{CH}_2$  group in the paraffin side chain of the amino acid diminishes this effect (that is,  $K_1/C$  is positive in sign and equal to 0.029), whereas each hydroxyl group increases this effect ( $K_1/C = -0.039$ ). The value of the limiting slopes in  $-(\log \gamma)/C$  in Table 3 is positive in sign for glycine, is diminished by each  $\text{CH}_2$  group, but increased by each hydroxyl group. If the influence of the  $\text{CH}_2$  group diminishes from approximately  $-0.10$  for its effect on  $-(\log \gamma)/C$  of a polar molecule without paraffin side chains to a value of  $-0.01$  for an additional  $\text{CH}_2$  group in a molecule already possessed of a paraffin side chain, the influence of the hydroxyl group would appear to diminish from an approximately equally large positive value, as in the case of serine, to a far smaller effect for hydroxyproline.

The influence of a group is, of course, not independent of its position in the molecule. The methyl group attached to the positively charged nitrogen, as in sarcosine, has an effect close to that in the paraffin side chain of its isomer, alanine, being estimated at  $-0.008$  for the former as compared to  $-0.005$  for the  $\alpha$ -amino acid<sup>14</sup>. Betaine, with three methyl groups attached to the positively charged nitrogen, has a value of  $-(\log \gamma)/C$  of  $-0.150$ . The influence of each non-polar group is thus far greater in this case than for its isomer, valine. We must thus consider not only the nature and number of polar and non-polar groups but also their position in the molecule.

#### Activity Coefficients of Asparagine as a Function of Concentration of Other Dipolar Ions

**Influence of CONH Groups.** Asparagine is an  $\alpha$ -amino acid of low solubility and with no exposed  $\text{CH}_2$  group. Moreover, the terminal group contains the CONH group, characteristic of the peptide linkage. Ammonia can be split from this group of asparagine and of the closely related glutamine in acid solution<sup>15, 16</sup>, and the concentration of these amides in systems containing other amino acids can thus readily be investigated either by titration or by Nessler determination of the liberated ammonia.

Most amino acids increase the solubility of asparagine, and the activity coefficients of the asparagine calculated from such measurements are given in Table 4. The solubility increased from 0.184 mole per liter in water to only 0.192 in 1.5 molal alanine and decreased to 0.170 mole per liter

TABLE 4. ACTIVITY COEFFICIENTS OF *l*-ASPARAGINE IN AQUEOUS AMINO ACID SOLUTIONS CALCULATED FROM SOLUBILITY MEASUREMENTS OF COHN, McMEEKIN, FERRY AND BLANCHARD<sup>13</sup> AT 25°

Amino acid concentration $C_2$ moles per liter	Density of solution $\rho$	Dielectric constant of solvent $D$	Solubility of <i>l</i> -asparagine		$-\log \gamma$ (observed)	$-\log \gamma$ (calculated)
			$C_3$ moles per liter	$N$ mole fraction		
<i>l</i> -Asparagine in water						
0.0	1.00714	78.5	0.184	0.00336		
<i>l</i> -Asparagine in glycine. $\delta = 22.6$ ; $K_R^* = 0.094$ ; $K_s^* = -0.006$						
0.25	1.01544	84.2	0.194	0.00357	0.026	0.023
0.50	1.02336	89.8	0.199	0.00370	0.042	0.044
1.00	1.03886	101.1	0.211	0.00397	0.073	0.079
1.50	1.05383	112.4	0.225	0.00430	0.107	0.108
2.00	1.06861	123.7	0.231	0.00449	0.126	0.132
2.80	1.09080	141.8	0.247	0.00494	0.167	0.163
<i>l</i> -Asparagine in diglycine. $\delta = 70.6$ ; $K_R^* = 0.136$ ; $K_s^* = -0.029$						
0.25	1.02164	96.2	0.196	0.00364	0.035	0.035
0.50	1.03508	113.8	0.201	0.00380	0.054	0.061
1.00	1.06120	149.1	0.218	0.00425	0.102	0.101
1.40	1.08170	177.4	0.224	0.00449	0.126	0.125
<i>l</i> -Asparagine in lysylglutamic acid. $\delta = 345$ ; $K_R^* = 0.150$ ; $K_s^* = -0.075$						
0.098	1.01762	112.4	0.186	0.00346	0.018	0.018
0.192	1.02780	144.8	0.189	0.00356	0.030	0.030
<i>l</i> -Asparagine in alanine. $\delta = 22.6$ ; $K_R^* = 0.094$ ; $K_s^* = 0.033$						
0.25*	1.01436	84.2	0.188	0.00347	0.014	0.014
0.50	1.02132	89.8	0.191	0.00357	0.026	0.025
1.00	1.03490	101.1	0.192	0.00368	0.040	0.040
1.50	1.04788	112.4	0.192	0.00375	0.048	0.049
<i>l</i> -Asparagine in $\alpha$ -aminobutyric acid. $\delta = 22.6$ ; $K_R^* = 0.026$ ; $K_s^* = 0.012$						
0.25	1.01382	84.2	0.182	0.00338	0.003	0.003
0.50	1.02025	89.8	0.180	0.00339	0.004	0.005
1.00	1.03347	111.1	0.178	0.00346	0.013	0.008
1.50	1.04591	112.4	0.170	0.00342	0.008	0.009

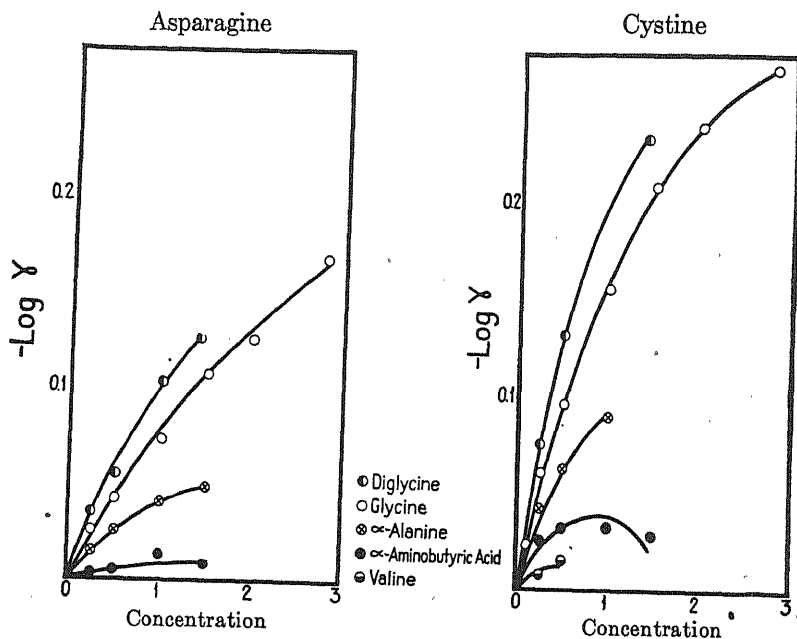
From Cohn, E. J., McMeekin, T. L., Ferry, J. D., and Blanchard, M. H., Ref. 13.

in 1.5 molal  $\alpha$ -aminobutyric acid. The greatest change in solubility produced by these two amino acids thus never exceeded 5 per cent, whereas



in Fig. 2a. The effects of the same substances on cystine are shown in Fig. 2b.<sup>17,18</sup>

At low concentrations the interaction of glycine with glycine and of glycine with asparagine would appear to yield almost identical activity coefficients.\* The large size of the asparagine molecule does not appreciably diminish this effect, which might thus be thought to depend largely on the moments of these two dipolar ions. Both are  $\alpha$ -amino acids, and the dielectric increment of glycine is reported to be somewhat larger than that of asparagine, being 22.6 for glycine<sup>19</sup> and 20.4 for asparagine<sup>20</sup> (see



Figures 2a and 2b. Solubility of Asparagine and of Cystine in solutions containing other Dipolar Ions.

Chapter 6). This difference would appear to be greater than the experimental error and might be attributable to the orientation of the terminal amide group with respect to the dipolar ion moment. The influence of glycine in diminishing the activity coefficient of asparagine is, however, not smaller but somewhat greater than the comparable effect of glycine upon glycine.

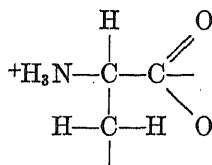
\* In other respects also the physical chemical behavior of glycine and asparagine are similar. Thus the change in free energy with change in solvent from water to ethanol at 25° of these two dipolar ions<sup>18</sup> is almost identical: the values of  $\log N/N_0$  for glycine and asparagine are 0.001 and 0.002, respectively.

**Influence of CH<sub>2</sub> Groups.** Terminal non-polar groups on interacting molecules have the opposite effect to that due to their electric moments. Thus alanine,  $\alpha$ -aminobutyric acid, and valine have activity coefficients greater than unity in aqueous solution. On the other hand, alanine and  $\alpha$ -aminobutyric acid decrease the activity coefficient of asparagine, although far less than glycine does. The apparent molal volume of  $\alpha$ -aminoisobutyric acid is 78.1 cc., or almost identical with that of asparagine, 78.0 cc., and that of valine is still greater, 91.3 cc.<sup>7</sup> Regardless of size, all the molecules with methyl groups have activity coefficients greater than unity in their aqueous solutions—and presumably also in the presence of each other—whereas the activity coefficients of asparagine are less than unity in alanine solutions.

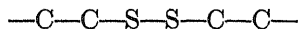
As a first approximation the CH<sub>2</sub> group possessed by alanine but not by glycine diminishes the value of  $-(\log \gamma)/C$  of asparagine by approximately 0.04 and the additional CH<sub>2</sub> group of  $\alpha$ -aminobutyric acid by something over 0.04 with respect to alanine. The influence of the CH<sub>2</sub> group upon interactions with asparagine, with its amide group, is thus of the same order as upon interactions with serine and threonine with their hydroxyl groups.

#### Activity Coefficients of Cystine as a Function of Concentration of Other Dipolar Ions

Cystine contains two positively charged ammonium groups, each in the  $\alpha$ -position to a negatively charged dissociated carboxyl group. In other words, each molecule contains two groups, each identical with that found in alanine, asparagine, and other  $\alpha$ -amino acids, the two



groups being connected by sulfur-sulfur linkage. Assuming freedom to rotate between the



atoms which separate the two dipoles, the moment of each of which may be considered as approximately 15 Debye units, the moment of cystine would be 30 Debye units, were the dipoles parallel, or zero were they anti-parallel. From the interaction of cystine and neutral salts (18) it was concluded that the moment was somewhat smaller than if the dipoles were parallel, but somewhat greater than would be expected if they were anti-parallel.

moment could be deduced were the theory for the interaction of dipolar ions completely satisfactory.

**Influence of CH<sub>2</sub> Groups.** The activity coefficients of cystine have been studied by the solubility method in the presence of  $\alpha$ -amino acids (Table 4 and Fig. 2b). Here also the effect of increasing the length of the paraffin side chain is to diminish the solvent action of the  $\alpha$ -amino acid. Taking the limiting slope of  $-(\log \gamma)/C$  of cystine in glycine (Fig. 2b) as 0.23, in alanine (Fig. 2b) as 0.16, in  $\alpha$ -aminobutyric acid as 0.10, and in *dl*-valine as 0.06, there would appear to be an effect of CH<sub>2</sub> groups, in diminishing the solvent action of dipolar ions, of slightly less than 0.07 in  $-(\log \gamma)/C$ . This is slightly smaller than the difference between glycine and alanine. The change from 1 to 2 or from 2 to 3 CH<sub>2</sub> groups in the side chain is, however, greater for the interaction with cystine, as for that with asparagine, and for hydroxy-amino acids in solutions of one solute, than for aliphatic  $\alpha$ -amino acids in solutions of one solute. It is thus possible that in the interactions of large molecules with strongly polar groups a decrease in  $-(\log \gamma)/C$  of 0.10 may be expected for each non-polar CH<sub>2</sub> group in paraffin side chains, as contrasted with a value of 0.01 for interactions involving CH<sub>2</sub> groups in less polar molecules (Table 3 and Tables 8 and 9 in Chapter 9).

The logarithm of the activity coefficient of cystine in the presence of varying concentrations of glycine is graphically represented in Fig. 2b. Comparison of the curve for cystine with that for asparagine (Fig. 2a) or glycine in glycine solutions (Fig. 1) indicates that a large factor in determining interactions between dipolar ions, as between ions and dipolar ions, is the magnitude of their electric moments.

**Influence of Electric Moments.** Opposite in sign to the effect of paraffin side chains, in determining the interactions between dipolar ions in aqueous solution, is the influence of their electric moments. For cystine, as for asparagine, the solvent action is greater for diglycine than for glycine. For urea, with a moment of approximately 5.1 Debye units<sup>21</sup>, the very definite solvent action is less than for glycine. *As a first approximation the change in free energy of interacting dipolar ions without paraffin side chains would thus appear to be proportional to the first power of their moments.*

#### Activity Coefficients of Peptides as a Function of Concentration

The activity coefficients of certain peptides have recently been studied in aqueous solutions by vapor pressure measurements by Smith and Smith<sup>22</sup>. As in the case of the  $\alpha$ -amino acids investigated, the electric moments of certain members of this series can be considered identical and

the change in activity coefficient as due to the number and position of methyl groups substituted for hydrogen in the molecule. On the other hand, the moments of all the peptides are appreciably higher (see Table 3,

TABLE 5. ACTIVITY COEFFICIENTS OF CYSTINE IN AQUEOUS AMINO ACID SOLUTIONS CALCULATED FROM SOLUBILITY MEASUREMENTS OF COHN, McMEKIN, FERRY AND BLANCHARD<sup>13</sup> AT 25°

Amino acid concentration $C_2$ moles per liter	Density of solution $\rho$	Dielectric constant of solvent $D$	Solubility of cystine		$-\log \gamma$ (observed)	$-\log \gamma$ (calculated)
			$C_3$ moles per liter	$N$ mole fraction		
Cystine in water						
0.0	0.9972	78.5	0.000454	0.00000820		
Cystine in glycine. $\delta = 22.6$ ; $K_R = 0.293$ ; $K_s = 0.007$						
0.1	1.0020	80.8	0.000478	0.00000865	0.023	0.022
0.25	1.00512	84.2	0.000518	0.00000942	0.060	0.052
0.5	1.01279	89.8	0.000558	0.0000102	0.095	0.095
1.0	1.02801	101.1	0.000632	0.0000117	0.154	0.161
1.5	1.0422	112.4	0.000701	0.0000132	0.207	0.207
2.0	1.05755	123.7	0.000743	0.0000142	0.238	0.238
2.8	1.08033	141.8	0.000779	0.0000152	0.268	0.267
Cystine in diglycine. $\delta = 70.6$ ; $K_R = 0.380$ ; $K_s = 0.0$						
0.25	1.01097	96.2	0.000532	0.00000976	0.075	0.078
0.5	1.02446	113.8	0.000597	0.0000111	0.131	0.131
1.4	1.07164	197.4	0.000711	0.0000140	0.232	0.235
Cystine in urea.* $K_R = 0.085$ ; $K_s = 0.0$						
1.0	1.0133	81.3	0.000536	0.00000994	0.083	0.082
2.0	1.02852	84.0	0.000622	0.0000119	0.162	0.156
Cystine in <i>dl</i> -alanine. $\delta = 22.6$ ; $K_R = 0.293$ ; $K_s = 0.136$						
0.25	1.00431	84.2	0.000495	0.00000904	0.042	0.033
0.5	1.01139	89.8	0.000513	0.00000947	0.062	0.060
1.0	1.02529	101.1	0.000532	0.00001004	0.088	0.092
Cystine in $\alpha$ -aminobutyric acid. $\delta = 22.6$ ; $K_R = 0.293$ ; $K_s = 0.191$						
0.25	1.00408	84.2	0.000474	0.00000869	0.025	0.021
0.5	1.01082	89.8	0.000475	0.00000884	0.032	0.033
1.0	1.02435	101.1	0.000461	0.00000884	0.032	0.037
1.5	1.03764	112.4	0.000442	0.00000875	0.028	0.021
Cystine in <i>dl</i> -valine. $\delta = 22.6$ ; $K_R = 0.293$ ; $K_s = 0.234$						
0.25	1.00382	84.2	0.000454	0.00000835	0.008	0.010
0.5	1.01046	89.8	0.000453	0.00000849	0.015	0.011

\* Values for  $D$  are taken from Wyman<sup>22</sup>.

From Cohn, E. J., McMeekin, T. L., Ferry, J. D., and Blanchard, M. H., Ref. 13.

(Chapter 6) than those of  $\alpha$ -amino acids and so are the observed values

These results are given in Table 6 which may be contrasted with Fig. 2, in which activity coefficients of asparagine in the presence of other dipolar ions are graphically represented. Both series of measurements illustrate the antagonistic effects due on the one hand to the influence of the electric moment in diminishing activity coefficients, on the other hand to the  $\text{CH}_2$  groups in paraffin side chains in increasing activity coefficients. For any electric moment, when the number of non-polar side chains reaches a certain value, the value of  $-(\log \gamma)/C$  passes through a maximum, as is the case for the interaction between asparagine and  $\alpha$ -aminobutyric acid (Fig. 2), and of pure solutions of alanylalanine (Table 6). In the latter case, as in pure solutions of alanine, and of  $\alpha$ -amino acids of ethyl and longer paraffin side chains, the value of  $-(\log \gamma)/C$  may even become negative.

TABLE 6. ACTIVITY COEFFICIENTS OF AMINO ACIDS AND PEPTIDES OF VARYING DIPOLE MOMENT FROM THE MEASUREMENTS OF SMITH AND SMITH<sup>14, 23</sup>

Moles per 1000 g. water	$\beta$ -Alanine*	<i>dl</i> - $\beta$ -Aminobutyric acid*	<i>dl</i> - $\beta$ -Aminovaleric acid*	$\gamma$ -Aminobutyric acid*	<i>dl</i> - $\gamma$ -Aminovaleric acid*	$\epsilon$ -Aminocaproic acid*	Glycylglycine	Glycylalanine	Alanyl-glycine	Alanylalanine	Tri-glycine
<i>m</i>	$-\log \gamma$	$-\log \gamma$	$-\log \gamma$	$-\log \gamma$	$-\log \gamma$	$-\log \gamma$	$-\log \gamma$	$-\log \gamma$	$-\log \gamma$	$-\log \gamma$	$-\log \gamma$
0.1											0.041
0.2	0.003	-0.003	-0.007	0.007	0.000	0.013	0.040	0.029	0.032	0.008	0.070
0.3							0.056	0.040	0.043	0.009	0.095
0.5	0.005	-0.008	-0.019	0.015	-0.004	0.022	0.082	0.054	0.056	0.007	
0.7	0.006	-0.012	-0.027	0.017	-0.008	0.024	0.102	0.061	0.063	0.001	
1.0	0.005	-0.021	-0.040	0.018	-0.016	0.026	0.128	0.068	0.068	0.015	
1.2	0.004	-0.027	-0.048	0.016		0.016	0.141	0.073			
1.5	0.000	-0.037	-0.062	0.011	-0.033	-0.001	0.157				
1.7							0.164				
2.0	-0.007	-0.056	-0.086	-0.003	-0.054	-0.030					

\* Activity coefficients are reported for higher concentrations than 2.0 molal in the original communication.<sup>14</sup>

Table 6 illustrates these two effects, namely, the effect of dipolar ions varying in moment in increasing the values of  $-(\log \gamma)/C$  and of those of the same moment but varying numbers of  $\text{CH}_2$  groups in diminishing  $-(\log \gamma)/C$ . The first effect would appear to be magnified at low concentrations, the second "salting-out" effect at higher concentrations.

### Interactions Between Dipolar Ions

In the preceding chapter we conclude that the interactions between organic solvents and dipolar ions could not be satisfactorily accounted for in purely electrostatic terms by any theory yet suggested. The very strong effects due to interactions of the non-polar groups both of solvent

Dipolar ions may be considered as occupying a position, stressed in the introduction to this book, intermediate between that of ions and most uncharged organic molecules. It therefore seemed possible that the interactions between dipolar ions—especially those without paraffin side chains—would more readily reflect the expected electrostatic forces.

TABLE 7. INFLUENCE OF CH<sub>2</sub> GROUPS UPON THE INTERACTIONS BETWEEN DIPOLAR IONS

Dipolar ion	CH <sub>2</sub> Groups in paraffin side chains $n_{CH_2}$	Limiting slope $-\log \gamma/C$	$K_R^*$	$K_s^*$	$\frac{\Delta K^*}{\%}$
Glycine .....	0	0.096	0.092	-0.004	
<i>dl</i> - $\alpha$ -Alanine .....	1	-0.005	0.010	0.015	0.019
<i>dl</i> - $\alpha$ -Aminobutyric acid .....	2	-0.033	0.016	0.049	0.027
<i>dl</i> - $\alpha$ -Aminovaleric acid .....	3	-0.041	0.000	0.042	0.015
<i>dl</i> - $\alpha$ -Aminoisobutyric acid .....	2	-0.054			
<i>dl</i> -Valine .....	3	-0.064			
$\beta$ -Alanine .....	0	0.015	0.022	0.007	
$\beta$ -Aminobutyric acid .....	1	-0.011	0.024	0.035	0.028
$\beta$ -Aminovaleric acid .....	2	-0.035	0.012	0.047	0.020
$\gamma$ -Aminobutyric acid .....	0	0.043	0.046	0.003	
$\gamma$ -Aminovaleric acid .....	1	-0.001	0.018	0.019	0.016
$\epsilon$ -Aminocaproic acid .....	0	0.079	0.098	0.019	
<i>dl</i> -Serine .....	1	0.075			
<i>dl</i> -Threonine .....	2	0.027	-0.014	-0.013	
Asparagine in glycine .....	0	0.100	0.094	-0.006	
Asparagine in alanine .....	1	0.061	0.094	0.033	0.039
Asparagine in $\alpha$ -aminobutyric acid .....	2	0.014	0.026	0.012	0.009
Cystine in glycine .....	0	0.226	0.293	0.067	
Cystine in $\alpha$ -alanine .....	1	0.157	0.293	0.136	0.069
Cystine in $\alpha$ -aminobutyric acid .....	2	0.102	0.293	0.191	0.062
Cystine in valine .....	3	0.059	0.293	0.234	0.056
Sarcosine .....	1	-0.008	0.016	0.024	
Betaine .....	3	-0.150	0.042	0.192	0.084
Glycylglycine .....	0	0.241	0.227	-0.014	
Alanylglycine .....	1	0.208	0.293	0.085	0.099
Glycylalanine .....	1	0.189	0.243	0.054	0.068
Alanylalalanine .....	2	0.073	0.203	0.130	0.072

For amino acid and peptide solutions the dielectric constant is, as we have seen, linear in the concentration,  $C$ , expressed as moles per liter, the

If we multiply the quantity in the Born-Fajans equation ( $1/D - 1/D_0$ ) by  $-D_0^2/\delta$ , we have

$$-(D_0^2/\delta)(1/D - 1/D_0) = (D_0/D)(D - D_0)/\delta \quad (3)$$

For the interaction between dipolar ions it will be more convenient to substitute for  $(D - D_0)/\delta$  in the last equation the concentration  $C$ , in moles per liter of the dipolar ion. The change in free energy due to electrostatic forces can therefore be conveniently written for interactions involving dipolar ions, in the form,

$$-(\bar{F}_e - \bar{F}_e^0)/2.303kT = -\log \gamma = K_R^*(D_0/D)C - K_s^*C \quad (4)$$

in which the salting-out constant  $K_s^*$  is equal to  $-K_1/C$ , and  $K_R^* = K_2\delta/2.303D_0^2kT$ . On the basis of the above equation the logarithm of activity coefficients of dipolar ions should be linear in  $(D_0/D)C$  in cases where the salting-out term is small. This is nearly true for the interaction between asparagine and glycine (Table 4) and between cystine and diglycine (Table 5). In other cases, if this equation be valid and  $(\log N/N_0)/C$  be plotted as ordinate against  $D_0/D$  as abscissa, extrapolation to the point where  $D_0/D$  equals unity should yield  $(K_R^* - K_s^*)$ . The difference between this value and that where  $D_0/D$  equals 0.5 should yield  $K_R^*/2$ . In this way estimates of  $K_R^*$  and  $K_s^*$  for each interaction can be obtained, provided the experimental points fall on a straight line when plotted in this manner. Certain of the measurements reported are analyzed in this way in Table 7. The limiting value of the slope,  $-(\log \gamma)/C$ , at zero concentration gives  $K_R^* - K_s^*$ .

In these terms the activity coefficients of alanine,  $\alpha$ -aminobutyric acid, and  $\alpha$ -aminovaleric acid are most readily described in terms of large values of the salting-out constant  $K_s^*$ , and small values of  $K_R^*$ . *Large values of  $K_s^*$  have been observed in all interactions in which one or both dipolar ions have paraffin side chains*, and the values of  $K_s^*$  are, moreover, of the same order of magnitude as for the transfer to non-aqueous solvents.

In contrast to the influence of the paraffin side chain upon  $K_s^*$  is that of an amide or peptide group. We estimated that  $K_s^*$  was close to zero in the interaction between glycine and asparagine, and asparagine has an amide group.  $K_s^*$  has also been estimated to be zero for the interaction of diglycine and cystine, and diglycine contains the peptide linkage. In the interaction between asparagine and diglycine, or asparagine and lysylglutamic acid, *where amide and peptide groups are constituent parts of both interacting dipolar ions*,  $K_s^*$  appears to have a sign opposite to that observed for the other dipolar ions investigated. This is reminiscent of

ship between "salting-in" by certain electrolytes and the concentration has been observed.

The moments for glycine, diglycine and lysylglutamic acid are 15, 26, and 59 Debye units, respectively (Table 8). The observed limiting slopes for the interaction of asparagine with these three dipolar ions may be taken as 0.10, 0.17 and 0.23. Thus  $K_R^*$  increases by less than the first power of the moment. The interaction of cystine has been studied with urea, glycine, and diglycine, and their moments may be taken as 5.1, 15, and 26, or roughly as 1 is to 3 is to 5. The observed limiting slope for cystine in urea, 0.088, is less than a third that in glycine, and that in diglycine, 0.380, is less than five times that in urea. As in the case of interactions with asparagine, *interactions between the dipolar ions that*

TABLE 8. INFLUENCE OF DIPOLE MOMENT UPON THE INTERACTIONS BETWEEN DIPOLAR IONS WITHOUT PARAFFIN SIDE CHAINS\*

Dipolar ions		Distance of closest approach $b_i + b_k$	Dipole moments		Product of dipole moments $\mu_i \mu_k$	Limiting slope	
			$\mu_i$	$\mu_k$		Estimated from experiment $-(\log \gamma)/C$	Calculated from theory $-(\log \gamma)/C$
Cystine	Urea	6.54	27	5	135	0.088	0.012
Glycine	Glycine	5.64	15	15	225	0.096	0.053
Asparagine	Glycine	6.13	15	15	225	0.100	0.042
Asparagine	Diglycine	6.65	15	26	390	0.165	0.097
Cystine	Glycine	6.76	27	15	405	0.226	0.101
Asparagine	Lysylglutamic acid	7.68	15	59	885	0.225	0.343
Diglycine	Diglycine	6.64	26	26	676	0.241	
Cystine	Diglycine	7.28	27	26	702	0.380	0.246
Triglycine	Triglycine		32	32	1024	0.470†	

\* The measurements upon such amino acids as  $\beta$ -alanine,  $\gamma$ -aminobutyric and  $\epsilon$ -aminocaproic acids are not here considered because of the lack of reliable measurements upon such molecules. (See Cohn, E. J., McMeekin, T. L., and Cohn, E. J., *J. Am. Chem. Soc.*, **56**, 2270 (1934); and McMeekin, T. L., Cohn, E. J., and Cohn, E. J., *J. Am. Chem. Soc.*, **56**, 2270 (1934).)

† For triglycine  $K_R^*$  has been estimated from the limiting slope.

have been investigated indicate that change in free energy with change in moment increases by slightly less than the first power of the products of the moments (Table 7).

**Theoretical Models.** Scatchard and Kirkwood have extended the treatment for the change in free energy with change in dielectric constant of the solvent to the case of dipolar ions<sup>25</sup>. Considering a dipolar ion as made up of two spheres of radius  $b$  separated by a distance  $R$ , with a charge  $+ez$  in one sphere and  $-ez$  in the other, they evaluate  $(\bar{F}_e - \bar{F}_e^0)/(1/D - 1/D_0)$  as  $N\epsilon^2 z^2 (1/b - 1/R)$ .

If we assume the last expression gives  $K_R$  (see equation 4) or  $2.022D^2$



then when the solubility of a given dipolar ion is influenced by different substances,  $K'_R$  should increase directly with the dielectric constant increment,  $\delta$ .

In order to test this equation we have tentatively assumed that  $R$  for glycine and all  $\alpha$ -amino acids is 3.17 Å. Considering glycine to be a sphere of radius 2.82 Å, with the center of the dipole at the center of the molecule, the charges would be 1.24 Å from the edge of the molecule. Taking  $b$  as 1.24 Å throughout, values of  $K'_R$  estimated on this double sphere or "dumb-bell model" are given in Table 9.

This model does not allow for the volume occupied by parts of the molecule not situated between the charged groups. Kirkwood has developed a spherical model<sup>26</sup> within which any number of charges is located. The general expression for the change in free energy with the dielectric constant of the medium is in the form of an infinite series. However, for the special case of two charges equidistant from the center, an explicit summation has been made. Taking the values of the radius  $b$ , given in column 3 of Table 9, and assuming the distance of the charges from the edge to be 1.24 Å, as before, values of  $K'_R$  have been calculated\* and listed in column 7. Here too  $K'_R$  should, according to the theory, be proportional to  $\delta$ .

In both these theoretical treatments the dipole molecules whose activity coefficients are considered are supposed to be surrounded by a structureless dielectric continuum, in which the presence of other dipolar ions increases the dielectric constant, their molecular structure being ignored. This picture may be replaced by one somewhat more detailed, which considers molecular interaction of pairs of dipolar ions immersed in a supposedly continuous medium with the dielectric constant of the pure solvent. Here the greater complexity of the treatment has led to a simpler model for the dipolar ion. In the development of Fuoss<sup>28, 29</sup> the molecule is represented by a sphere of radius  $b$  with a point dipole at the center.† Following Fuoss, the interaction constant of a dipole species  $i$  in the presence of an excess of a species  $k$  in water at 25° is given by

$$K_R = -(\log \gamma_i)/C_k = 3.69 \times 10^{-4} \mu_i \mu_k \theta(x) \quad (5)$$

where  $\mu_i$  and  $\mu_k$  are the moments of the two dipoles in Debye units and  $\theta(x)$  is a function given by Fuoss<sup>29, Table 2</sup>, where

$$x = 0.706 \mu_i \mu_k / (b_i + b_k)^3 \quad (6)$$

\* Kirkwood, J. G., *Chem. Rev.*, 19, 275 (1938).

† For hemoglobin the two charges were taken as  $+2e$  and  $-2e$ , respectively, consistent with the dipole moment of 500<sup>27</sup> and  $R = 51.8$  Å. (See Ferry, Cohn and Newman: *J. Am. Chem. Soc.*, 60, 1480 (1938)). See Chapter 9.

TABLE 9. CONSTANTS FOR THE INTERACTION OF DIPOLAR IONS

(1)	(2)	(3)	(4)	(5)	(6)		(7)	(8)	(9)	(10)	(11)	(12)	(13)
Dipolar ion	Molar volume of dipolar ion	Radius of dipolar ion	Dipole moment of dipolar ion	Distance of closest approach	Constant calculated from dependence on dielectric constant of medium		Sphere model	Constant calculated from dipole-dipole interaction	Calculated salting-out constant	Calculated limiting slope	Observed limiting slope	$\frac{D_0}{D} \frac{K_R - K_S}{K_R}$	Estimated from $-\log \gamma / C^*$
					Dumb-bell model	$K'_R$							
Interactions with one solute													
Interactions with asparagine ( $\mu = 15$ ; $b = 3.31 \text{ \AA}$ .)													
.....	57.0	2.82	15	5.64	0.43	0.35	0.053	0.022	0.022	0.031	0.098	0.100	0.007
ine.....	73.3	3.08	15	6.16	0.43	0.35	0.041	0.022	0.022	0.019	-0.011		0.011
isobutyric acid.....	89.6	3.29	15	6.58	0.43	0.34	0.033	0.022	0.022	0.011	-0.043		0.043
isovaleric acid.....	105.9	3.46	15	6.92	0.43	0.34	0.029	0.022	0.022	0.007	-0.047		0.047
Interactions with cystine ( $\mu = 27$ ; $b = 3.94 \text{ \AA}$ .)													
.....	57.0	2.82	15	6.13	0.43	0.34	0.042	0.021	0.021	0.021	0.100	0.094	-0.006
ine.....	93.3	3.34	26	6.65	1.36	1.06	0.097	0.043	0.043	0.054	0.165	0.136	0.029
lutamic acid.....	211.5	4.37	59	7.68	6.63	5.20	0.343	0.122	0.122	0.221	0.225	0.150	-0.075
ine.....	73.3	3.08	15	7.02	0.43	0.34	0.037	0.022	0.022	0.015	0.061	0.094	0.033
isobutyric acid.....	89.6	3.29	15	6.60	0.43	0.34	0.033	0.022	0.022	0.011	0.014	0.026	0.012
Interactions with hemoglobin ( $\mu = 500$ ; $b = 27 \text{ \AA}$ .)													
.....	44.0	2.59	5	6.54	0.07		0.012	0.019	0.019	-0.007	0.088	0.293	0.067
e.....	57.0	2.82	15	6.76	0.56		0.101	0.037	0.037	0.064	0.226	0.380	0.000
ine.....	93.3	3.34	26	7.28	1.75		0.246	0.068	0.068	0.178	0.380	0.293	0.000
ine.....	73.3	3.08	15	0.56	0.56		0.090	0.039	0.039	0.051	0.157	0.293	0.136
isobutyric acid.....	89.6	3.29	15	7.23	0.56		0.081	0.040	0.040	0.041	0.102	0.293	0.191
.....	105.9	3.46	15	7.40	0.56		0.076	0.041	0.041	0.035	0.059	0.293	0.234
Interaction with hemoglobin ( $\mu = 500$ ; $b = 27 \text{ \AA}$ .)													
e.....	57.0	2.82	15	30		2.67	0.39	0.14	0.14	0.25	0.95	1.90	0.95

culated from the assumption of spherical shape.

m Cohn, E. J., McMeekin, T. L., Ferry, J. D., and Blanchard, M. H., Ref. 13.

and  $b_i$  and  $b_k$  are the radii of the two dipoles. Here both  $x$  and  $K_R$ , as given by Fuoss, have been multiplied by the factor 9/4, in accordance with the correction of Kirkwood<sup>9, 30</sup>, for the case that the dielectric constant of the cavities represented by the molecules is small compared with that of the surrounding solvent.

Dipole interaction has been calculated by means of the above equation for the molecules that have been investigated, using the values of  $\mu$  and  $b$  given in Tables 8 and 9. These values of  $K_R$  are listed in the last column of Table 8.

In order to account for the salting-out effect in dipole-dipole interaction Kirkwood<sup>9</sup> has suggested for the same spherical model an equation of the form

$$K_s = \frac{3\pi N}{2303DkT} \frac{1}{2(b_i + b_k)^3} (\mu_i^2 b_k^3 + \mu_k^2 b_i^3) \quad (7)$$

with the units in electrostatic units which for water at 25° with  $\mu$  in Debye units becomes:

$$K_s = [3.82 \times 10^{-4} / (b_i + b_k)^3] [\mu_i^2 b_k^3 + \mu_k^2 b_i^3] \quad (8)$$

Values of  $K_s$  calculated on the basis of this equation have been given by Cohn, McMeekin, Ferry and Blanchard\* (Column 9 in Table 9).

These various calculations have been compared with the experimental limiting slopes<sup>13</sup>. The only cases in which there is even approximate agreement with the Born-Fajans treatment are those of cystine in urea and hemoglobin in glycine. In each of these systems the molecule whose solubility is measured is large compared with the other dipolar ionic species, so that the merging of the latter in a continuous medium, assumed by the theory, is more closely approached. In all other systems investigated the calculated values are many times larger than those observed, the discrepancy being greater for the dumb-bell than for the spherical model.

Corrected or uncorrected for the salting-out effect, the model for interaction between spherical molecules with dipoles at their centers gives far smaller results, more nearly of the order of the experimentally observed limiting slopes (Table 8, last column). Two types of discrepancy are apparent. For the activity coefficients of glycine, asparagine, and cystine in glycine, diglycine and lysylglutamic acid, the calculated limiting slope is too small; presumably the dipoles are mutually accessible to a greater extent than represented by the model. On the

\* Reference 13, Table 4. An alternative molecular model suggested by Kirkwood<sup>9</sup> for calculations of the interaction constant  $K_R$  for water

other hand, for the activity coefficients of  $\alpha$ -amino acids with paraffin side chains and for asparagine in  $\alpha$ -aminobutyric acid, the calculated limiting slope is too large; presumably the repulsive effect of the side chains is greater than the model represents.

While the agreement of the different theoretical approaches with experiment is nowhere satisfactory, the nature of the discrepancies suggests that a much better representation of the facts may be obtained by extending the treatment of dipole-dipole interaction to molecular models of a more detailed and specific structure.

## Chapter 11

### Interactions between Ions and Amino Acids or Peptides

BY EDWIN J. COHN

The interactions of proteins and electrolytes had been observed in the century before the interactions of amino acids and electrolytes were investigated. Denis<sup>1</sup> noted that certain of the proteins of the blood, later called globulins, were soluble in neutral salt solutions but not in salt-free solutions. In 1905 there appeared three classical studies upon globulins with special reference to their interactions with electrolytes<sup>2, 3, 4</sup>. In one of these, upon serum globulin, Mellanby concluded: "Solution of globulin by a neutral salt is due to forces exerted by its free ions. Ions with equal valencies, whether positive or negative, are equally efficient, and the efficiencies of ions of different valencies are directly proportional to the squares of their valencies"<sup>3, p. 375</sup>. This accurate formulation of the principle of the ionic strength—rediscovered in 1921 by G. N. Lewis<sup>5</sup> as a description of the effect of neutral salts upon each other—acquired theoretical significance in Debye's theory of interionic forces. The ionic concentration,  $\Gamma$ , defined as the summation  $\sum Cz^2$ , where  $z$  is the valence of each ionic species, is related in Debye's theory to the temperature, the dielectric constant of the medium and  $\kappa$ , a measure of the thickness of the ion atmosphere, by the equation

$$\kappa^2 = \frac{4\pi N\epsilon^2}{1000DkT} \sum Cz^2 = \frac{12.67 \times 10^{13}}{DT} \Gamma \quad (1)$$

The dielectric constant thus plays an important role in this theory in determining that part of the activity coefficients of ions due to Coulomb forces,  $\gamma$ .

$$-\log \gamma = \left( \frac{\epsilon^2 z^2}{2.303 \times 2DkT} \right) \left( \frac{\kappa}{1 + \kappa a} \right) \quad (2)$$

In this equation the dielectric constant is considered unaffected by change in ionic strength. (This is comparable to equation 141 in Chapter 3).

parable to that employed by Debye and McAulay<sup>6</sup> and by Hückel<sup>7</sup> must also be added, (see equations 140 and 160 in Chapter 3), in which "salting-out" is considered some function of  $b$ , the ionic radii.

Since biological systems contain ions of various kinds as well as other components, some of which presumably decrease and others of which are known to increase the dielectric constant of water, the significance of these relations can scarcely be overemphasized. Were the interaction between dipolar ions and ions completely defined by the above equation, and were it necessary merely to substitute in equation 2 the change in the dielectric constants of solutions due to biological components in order to estimate the change in the activity coefficients of ions in biochemical systems, the problem would indeed be simple, at least in sufficiently dilute salt solutions. The influence of different ions would then depend upon their radii,  $b$ , their mean effective diameters,  $a$ , and valence,  $z$ , and the dielectric constant.

The Coulomb forces due to ions are not, however, the only ones which we must consider. The conditions in which the activity coefficients of ions and dipolar ions may be expected to depend most completely upon Coulomb forces obtain in media of low dielectric constant at low temperature. Conversely, the specific properties of ions and dipolar ions manifest themselves the more the higher the temperature and the dielectric constant.

In dilute solutions of electrolytes the logarithm of the activity coefficient is proportional to  $\kappa$  and to the square root of the ionic strength.

$$\kappa = 0.33 \times 10^8 \sqrt{(D_0/D)(\Gamma/2)} \quad (3)$$

In the case of dipolar ions the logarithm of the activity coefficient does not vary as the square root, but as the first power of the concentration. This was empirically discovered by studying the solvent action of neutral salts upon cystine, and theoretically demonstrated by Scatchard and Kirkwood<sup>8</sup> in an extension of the theory of Debye and Hückel to dipolar ions. They demonstrated that the term proportional to the square root of the concentration vanishes when the net charge is zero.

### The Solubility of Dipolar Ions in Aqueous Salt Solutions

The influence of neutral salts upon the solubility of amino acids was studied by Pfeiffer and his co-workers<sup>9, 10</sup> almost 80 years after the discovery of the globulins by Denis. Certain of their results are graphically represented in Fig. 1, in which the logarithm of the ratio of solubility of the amino acid in the salt solution to that in water is given as ordinate and

<sup>6</sup> Debye, P., and McAulay, J., *Physik. Z.*, 26, 22 (1925).

the concentration, in this case equivalent to the ionic strength, as the abscissa.

Pfeiffer and Würigler thus noted the very great difference in the influence of sodium and potassium chloride in their interactions respectively with glycine and leucine. For whereas glycine was dissolved by sodium chloride and by small concentrations of potassium chloride, (higher concentrations of the latter salt slightly diminished solubility) the solubility of leucine

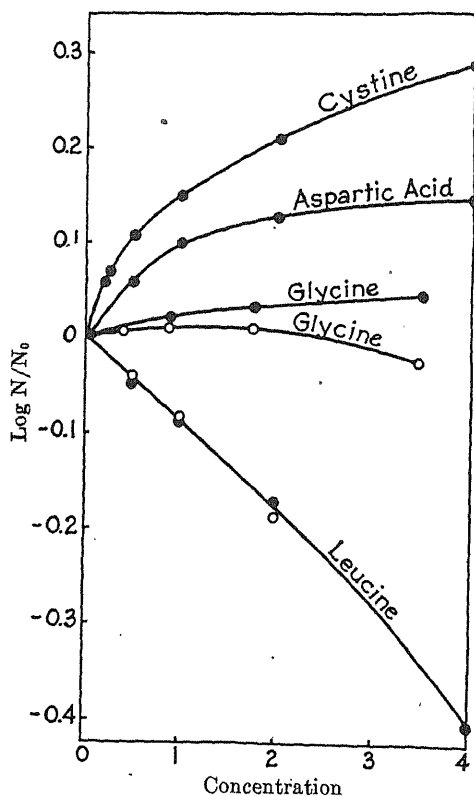


Figure 1. Solubility of amino acids in NaCl, ●, and KCl, ○.

was diminished by both sodium and potassium chloride, salting-out being the predominant effect.

Euler and Rudberg<sup>11</sup> observed an even greater salting-out effect for tyrosine by sodium chloride and suggested that this phenomenon could be used in the separation of amino acids much as it is for proteins.

Glycine and leucine, as we have seen, possess the same polar groups

groups and in the size of the molecule. They also differ in the far greater solubility of glycine in water, the latter depending, crystal lattice energies being equal, upon repulsive forces between the polar solvent and the non-polar paraffin side chain of the dipolar ions (compare with Chapter 10). That these non-polar groups of leucine are responsible for its being readily salted-out by sodium chloride, rather than its larger size or lower solubility, may be demonstrated by comparing the behavior of leucine with that of aspartic acid (Fig. 1). Aspartic acid is less soluble than leucine and its molecular size is far greater than that of glycine. Sodium chloride has a far greater solvent action upon aspartic acid than upon either glycine or leucine. That this is not due to the second carboxyl group of aspartic acid but to the absence of a paraffin side chain terminating in a methyl group may be demonstrated by comparing these results upon aspartic acid with those more recently obtained with asparagine (Table 1).

Different electrolytes have very different solvent actions upon amino acids as upon proteins. The influence of calcium chloride upon the same amino acids is graphically represented in Fig. 2. The order in which solubility is increased by calcium chloride is the same as for sodium chloride. The solvent action of calcium chloride is, however, far greater. Moreover, even leucine, which is salted-out by sodium or potassium chloride, is dissolved by calcium chloride. The latter would thus appear to have a far smaller salting-out effect. Empirical analysis of the curves in Fig. 2 might suggest that the solvent action of calcium chloride of low concentration was the same for glycine and for leucine. This approximates the truth. The very profound effect of calcium chloride in diminishing activity coefficients, graphically represented in Fig. 2, demands further analysis, however, for the interactions between glycine and calcium chloride in water are still greater at lower glycine concentration, approaching those of aspartic acid.

**Solubility of Asparagine in Aqueous Salt Solutions.** The simplest of the  $\alpha$ -amino acids, glycine, is, as we have seen, so soluble that the influence of dipole-dipole interactions cannot be neglected, and it was thought possible that these might account for the far greater change of solubility with ionic strength of aspartic acid than of glycine (Fig. 1). Aspartic acid on the other hand, with its two carboxyl groups, is an acid molecule at its isoelectric point. Asparagine is closely related to aspartic acid, is neutral, and has been found to behave much as does glycine in dipole-ethanol (Chapter 9) and dipole-dipole interactions (Chapter 10) despite its larger size.

The solubility ratios of asparagine in aqueous sodium and lithium



same and are far greater than upon glycine. The measurements upon glycine in lithium chloride carried out at 25° (reported in Table 13), are in fair agreement with those of Pfeiffer and Würzler (10) at 20°, which they supplement, and allow us to carry still further the analysis of the interactions of  $\alpha$ -amino acids and alkali halides.

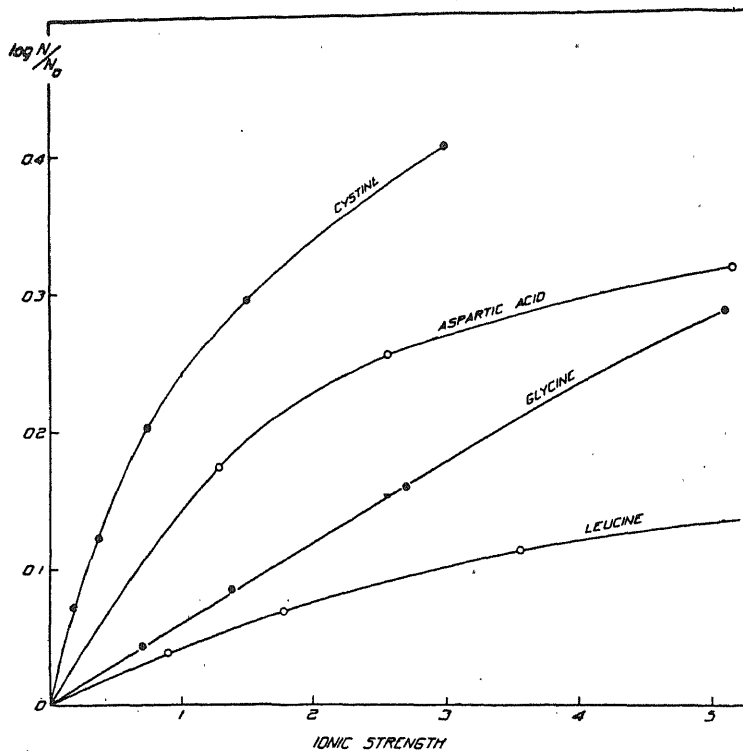


Figure 2. Solubility of amino acids in  $\text{CaCl}_2$ . The data for cystine are from the paper by Blix [*Z. physiol. Chem.*, 178, 109 (1928)]. The other data are reported by Pfeiffer and Würzler [*Z. physiol. Chem.*, 97, 128 (1916)]. From Edsall, J. T., in Schmidt, C. L. A., *The Physical Chemistry of the Amino Acids and Proteins*, p. 871, Thomas, Springfield, Ill., 1938.

The difference between the interactions of alkali halides with these  $\alpha$ -amino acids would appear to depend rather upon the greater solubility of the glycine than upon the nature of alkali halide. The measurements upon the solubility ratio of asparagine in solutions containing both sodium chloride and glycine<sup>11a</sup>, also reported in Table 1, are graphically represented in Fig. 3. It will be noted that the higher the glycine concentration the

# INTERACTIONS BETWEEN IONS AND AMINO ACIDS 241

smaller the solvent action of the sodium chloride upon the asparagine. The influence of dipole-dipole interaction in decreasing the interaction of ions and dipolar ions is thus demonstrated by these measurements.

TABLE 1. INTERACTIONS OF ASPARAGINE WITH DIPOLAR IONS AND SALTS AT 25°

Ionic strength of electrolyte $I/2$	Density of solution $\rho$	Amino acid concentration	Solubility		$\left(\frac{D_0}{D} \frac{I}{2}\right)$	Logarithm of solubility ratio $\log N/N'$	$\left(\frac{D}{D_0} \log \frac{N}{N'}\right)$
			Moles per liter $C_s$	Mole fraction $N$			
Asparagine in Water							
0.0	1.00714	0.0	0.184	0.00336	0.0	0.0	0.0
Asparagine in NaCl							
0.25	1.01821		0.203	0.00371	0.236	0.043	0.045
0.50	1.02848		0.216	0.00396	0.471	0.071	0.075
1.00	1.04881		0.241	0.00442	0.935	0.119	0.127
2.00	1.08782		0.290	0.00536	1.846	0.203	0.220
Asparagine in LiCl							
0.10	1.01006	0.0	0.193	0.00353	0.095	0.0215	0.023
0.25	1.01409		0.203	0.00371	0.236	0.0431	0.046
0.50	1.02055		0.216	0.00395	0.471	0.0703	0.075
0.75	1.02698		0.229	0.00420	0.704	0.0969	0.103
Asparagine in Glycine and NaCl							
0.00	1.03886	1.0	0.211	0.00397	0.0	0.0	0.0
0.50	1.05890	1.0	0.239	0.00451	0.369	0.055	0.075
1.00	1.07816	1.0	0.259	0.00490	0.734	0.091	0.124
2.00	1.11556	1.0	0.298	0.00569	1.456	0.156	0.214
0.00	1.06861	2.0	0.231	0.00449	0.0	0.0	0.0
0.50	1.08758	2.0	0.248	0.00484	0.304	0.033	0.054
1.00	1.10600	2.0	0.266	0.00520	0.605	0.064	0.106
Asparagine in $\alpha$ -Amino- <i>n</i> -butyric Acid and NaCl							
0.0	1.02025	0.5	0.180	0.00339	0.0	0.0	0.0
0.1	1.02461	0.5	0.189	0.00356	0.084	0.0212	0.025
0.5	1.04086	0.5	0.208	0.00393	0.416	0.0642	0.077
1.0	1.06069	0.5	0.233	0.00441	0.826	0.1142	0.138
0.0	1.03347	1.0	0.178	0.00346	0.0	0.0	0.0
0.1	1.03750	1.0	0.183	0.00355	0.075	0.0111	0.015
0.5	1.05351	1.0	0.201	0.00391	0.372	0.0531	0.071
1.0	1.07300	1.0	0.223	0.00436	0.740	0.1004	0.136

Solubility measurements reveal the Coulombic interactions of ions and dipolar ions only when the solubility of the saturating body is very small.

amino acids. In order to estimate activity coefficients by the solubility method it seems desirable to consider this tetrapole which possesses no paraffin side chain and which has a solubility of only 0.019 gram per liter at 25°. The solubility measurements upon cystine may therefore be expected to yield activity coefficients without correction for dipole-dipole interactions.

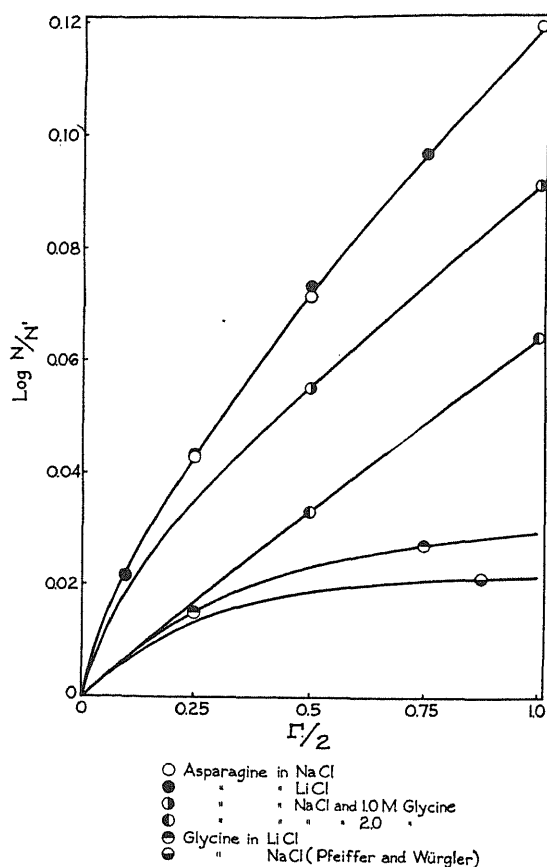


Figure 3. Solubility of glycine and asparagine in sodium chloride and lithium chloride.

The influence of calcium chloride upon the solubility of cystine has been studied by Blix<sup>12</sup>. His measurements (Fig. 2) have been confirmed and the influence upon cystine of a large number of other salts investigated<sup>13</sup> (Fig. 4). The fan-like spread of solubility in the presence of different

salts at the same ionic strength, multivalent cations having the greatest and multivalent anions the smallest solvent action, resembles that of certain proteins (see the solubility of carboxyhemoglobin in aqueous salt solution, Fig. 8, Chapter 24).

Although the solvent action of sodium chloride upon cystine is far smaller than that of calcium chloride (Fig. 4), that of ammonium sulfate and sodium sulfate is smaller still. The salting-out effect of sodium sulfate is greater than is that of ammonium sulfate, but could not be studied in such concentrated salt solution. In sufficiently concentrated salt solutions the precipitating action of ammonium sulfate could indeed

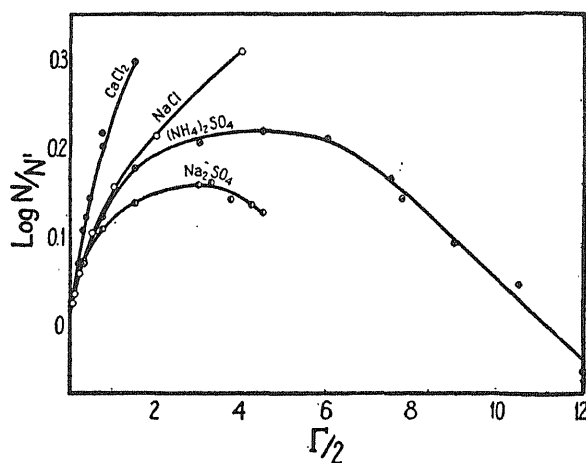


Figure 4.  
Solubility of cystine in  
aqueous salt solutions.

be defined by the salting-out expression which applies equally to gases, amino acids and proteins:

$$\log N = \beta' - K'_s(\Gamma/2) = \beta'' - K''_s I/2 \quad (4)$$

in which  $\Gamma/2$  is the ionic strength per liter, and  $I/2$  the ionic strength per 1000 grams of water. Taking the ionic strength range from 6 to 12, the solubility results of cystine in ammonium sulfate are analyzed in these terms in Table 2.

Any solubility equation for cystine, as for proteins, must therefore have the characteristic that  $\log N$  becomes linear in the ionic strength in sufficiently concentrated salt solutions. The value for both  $K'_s$  and  $K''_s$  and  $\beta'$  and  $\beta''$  are given in the last columns of Table 2. These results

of cystine in water ( $\log N' = -5.086$ ) is 0.396. An adequate expression for the activity coefficient of cystine in ammonium sulfate solutions more concentrated than two molal—but not in more dilute salt solution—would therefore be

$$-\log \gamma = \log N/N' = 0.396 - 0.027I/2 \quad (4a)$$

The large salting-out effect observed in these concentrated ammonium sulfate solutions might be considered as being manifest also at lower ionic strength. The principle of the ionic strength could be expected to obtain therefore only when the salting-out effect is small in comparison with Coulomb forces.

In the case of ions measurements at low concentrations have been employed in order to distinguish between Coulomb forces and the salting-out effect. For dipolar ions, both Coulomb forces and the salting-out

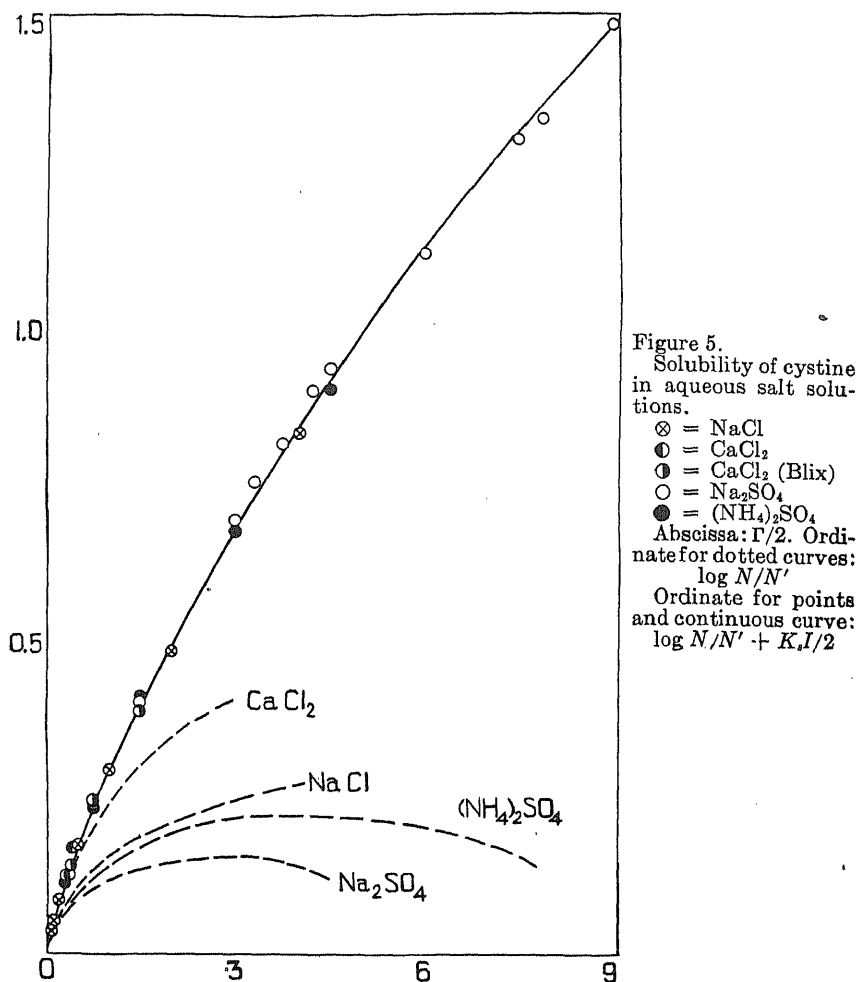
TABLE 2. SOLUBILITY OF CYSTINE IN CONCENTRATED AMMONIUM SULFATE SOLUTIONS DEFINED BY THE EQUATIONS:  $\log N = \beta' - K_s' (\Gamma/2) = \beta'' - K_s'' I/2$

Ionic Strength		Cystine Solubility mole fraction $N \times 10^3$	$\beta'$	$\beta''$	$K_s'$	$K_s''$
$\Gamma/2$	$I/2$					
6.00	6.915	1319	-4.66	-4.70	0.046	0.027
7.50	9.021	1190	-4.60	-4.69	0.043	0.026
7.89	9.630	1133	-4.60	-4.70	0.044	0.027
9.00	11.340	1010	-4.63	-4.70	0.044	0.027
10.50	13.965	920	-4.61	-4.67	0.041	0.025
12.00	16.848	726	-4.53	-4.70	0.045	0.027
Average.....			-4.61	-4.69	0.044	0.027

effect are linear in the ionic strength. We may therefore attempt an analysis of the results in Fig. 4 by assuming that, as a first approximation, differences in the solubility curves in the various solvents are completely ascribable to differences in the salting-out constant,  $K_s$ .

Provided  $K_s$  be assumed to be 0.06 for calcium chloride, 0.14 for sodium chloride, 0.152 for ammonium sulfate, and 0.18 for sodium sulfate, and we plot  $\log N/N' + K_s I/2$  against  $\Gamma/2$ , all the results with these different salts fall very closely upon the same curve (Fig. 5). This is true although sodium chloride and calcium chloride have a solvent action upon cystine even in concentrated aqueous salt solution, whereas ammonium sulfate, as we have seen, is a precipitant under these conditions. The uncorrected values of  $\log N/N'$  are given by the dotted lines. The limiting slope, determined by Coulomb forces alone, may be considered equal to that

The shape of the solid curve in Fig. 5 has the property of being linear at very low, and again at very high, values of the ionic strength. An



empirical expression on the basis of which the curve in Fig. 5 has been constructed may be written as follows:

$$\log \frac{N}{N'} + K_s \frac{I}{2} = \frac{\Gamma}{2} \left( \frac{K_R + A(\Gamma/2)}{1 + B(\Gamma/2)} \right) \quad (5)$$

same value of  $K_R$ , 0.56, is assumed tentatively for the action of all salts on cystine, though the value of the limiting slope may ultimately prove to be greater and slightly different for different salts. The values of  $K_s$  are those given above.<sup>15</sup> At the highest ionic strengths studied the curvature is slightly smaller than that given by the above equation and appears to have a nearly constant slope in such concentrated salt solutions that salting-out is the predominant factor and equation 4 is an adequate description of the data. The form of the above equation merely gives us a method of expressing the complicated relations that obtain between ions and dipolar ions, at least as a first approximation.

**Activity Coefficients of Cystine in Aqueous Salt Solutions Containing Other Amino Acids.** As a preliminary to the consideration of systems containing neutral salts and more than one protein<sup>15a</sup>, we may consider systems containing neutral salts and two amino acids. The results of the measurements<sup>16</sup> on systems containing cystine, glycine, and sodium chloride are graphically represented in the accompanying figures (Figs. 6a and 6b), in which the solubility of the cystine is always given as mole fraction,  $N$ . The lowest curve in Fig. 6a gives the interaction between cystine and glycine in the absence of salt. In the systems containing glycine solubility is always higher, and higher the greater the concentration of glycine and of salt. Whereas both glycine and sodium chloride have a solvent action upon cystine, the influence of each of these solvents is to decrease the effect of the other. We must therefore conclude that *the interaction between these dipolar ions is a function of the sodium chloride concentration, decreasing with increasing ionic strength.*

The converse to the above statement regarding the influence of ionic strength on the interaction between dipolar ions may also be made, namely, that *the interaction between ions and the dipolar ion cystine is a function of the other dipolar ions in solution, diminishing with increasing glycine concentration.* The very profound solvent action of sodium chloride on cystine is greatly reduced in the presence of even a mole of glycine, and is almost negligible in the presence of a 2.8 molal glycine solution. In so far as the interaction between the sodium and chlorine ions and the charged groups of the tetrapole cystine depends upon electrostatic forces, this effect is understandable, for by definition change in free energy due to electrostatic forces must, according to Coulomb's law, diminish inversely

<sup>15</sup> At sufficiently high concentrations the terms in  $A$  and  $B$  will be negligible in comparison with the first term both of the numerator and of the denominator of the fraction. Therefore, the ratio  $A/B$  will be independent of concentration. At high ionic strengths, therefore, the ratio  $A/B$  will be a constant. At ionic strengths of 4.5, by a value for  $A$  of 0.38, the first term of the numerator and denominator of the fraction is negligible in comparison with the second term. If  $K_s'$  is multiplied by  $1/2$ , still higher ionic strengths are described by the equation.  $A/B$  must be somewhat greater and  $B$  may be taken as 2.0. The solid curve in Fig. 5 is constructed on this basis.

as the second power of the dielectric constant, and the dielectric constant of a 2.8 molal glycine solution is 141.8, or as much above water as water is above butanol.

The dielectric constant enters into Debye's electrostatic force theory in several ways. If we plot the measurements, made at different glycine concentrations and therefore at different dielectric constants, with  $(D_0/D)(\Gamma/2)$  (which is proportional to  $\kappa^2$ ) as abscissa and  $(D/D_0) \log N/N' + K_s\Gamma/2$  as ordinate, all the curves should coincide provided the interactions between ions and dipolar ions at such high dielectric constants are completely ascribable to Coulomb forces (see Chapter 12). So plotted these results are given in Fig. 7.

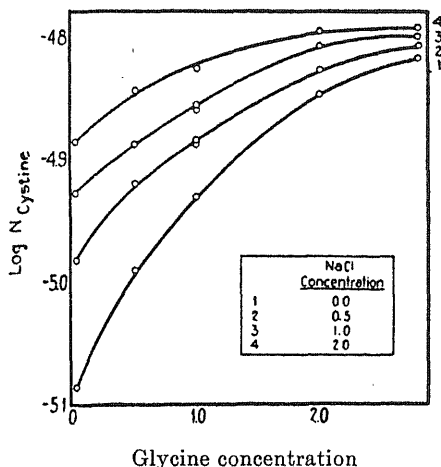


Figure 6a

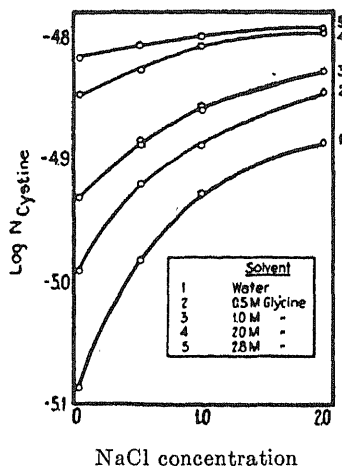


Figure 6b

Figure 6a. Influence of sodium chloride on the interaction between glycine and cystine.

Figure 6b. Influence of glycine on the interaction between sodium chloride and cystine.

Cystine exhibits both the solvent and salting-out effects characteristic of salts and of proteins. Extending equation 5 for interactions at dielectric constant,  $D$ , other than those of water at the same temperature,  $D_0$ , we have

$$\frac{D}{D_0} \log \frac{N}{N'} + K_s \frac{\Gamma}{2} = \frac{D_0 \Gamma K_R + A(D_0/D)(\Gamma/2)}{D \frac{\Gamma}{2} 1 + B(D_0/D)(\Gamma/2)} \quad (6)$$

This is the equation<sup>17</sup> on the basis of which the curve in Fig. 7 is drawn.

<sup>17</sup> In equation 6,  $N$  is the observed solubility and  $N'$  that at the same glycine concentration but zero ionic



More important than this semi-empirical equation is the fact that all the measurements upon cystine in different salt solutions and different concentrations of glycine fall on the same curve when plotted as in Fig. 7, and that the constants  $K_R$  and  $K_s$  have the values previously deduced from studies in aqueous solutions.

Kirkwood has suggested that the magnitude of the salting-out effect should be determined entirely by the relation

$$K_s = \frac{4\pi N \epsilon^2}{2303 D_0 k T} \frac{b^3}{a} \frac{D - 1}{2D + 1} \quad (7)$$

provided it depended upon the dipolar ion displacing a certain quantity of solvent and reducing the polarization of the solvent by the salt ion. In this expression  $b$  is the radius of the dipolar ions and  $a$  the sum of the radii of ions and dipolar ions. Taking  $b$  for glycine as  $2.82 \text{ \AA}^{17a}$  and  $a$  for cystine

which the dielectric constant is a function of the concentration of ions and dipolar ions, the absence of ions plus their effect on the dielectric constant is the limiting law for small salt concentrations to the salt concentration.

With these assumptions we may write the free energy of a solution containing ions and dipolar ions, at constant temperature and pressure, as

$$\begin{aligned} F/2.3RT = & n_1(F_1/2.3RT + \log N_1) + n_2(F_2/2.3RT + \log N_2) \\ & + n_3(F_3/2.3RT + \log N_3) - (D_0/D)^2 K_R n_2(\Gamma/2) \\ & + (D_0/D) K_s n_2(\Gamma/2) - (D_0/D)^2 K_R^* n_2 C_2/2 + (D_0/D) K_s^* n_2 C_2/2 \end{aligned}$$

in which  $n_1 F_1$  is the contribution to the non-ideal free energy of the solvent without ions or dipolar ions,  $n_2 F_2$  is the contribution of the dipolar ions at zero concentrations of dipolar ions and salt, and  $F_3$  is the contribution of the ions at zero concentration of dipolar ions.  $C_2$  is the concentration of dipolar ions,  $\Gamma$  is the dielectric constant of the solvent,  $K_R$  and  $K_R^*$  are the dielectric constants of the solvent and dipolar ions, respectively, and  $K_s$  and  $K_s^*$  are the dielectric constants of the solvent and dipolar ions, respectively. Then

$$\begin{aligned} \log a_2 = & \left[ \partial \left( \frac{F}{2.3RT} \right) / \partial n_2 \right]_{n_1, \Gamma} = F_2/2.3RT + \log N_2 \\ & - \left( K_R - \frac{D}{D_0} K_s \right) \left( \frac{D_0}{D} \right)^2 \frac{\Gamma}{2} - \left( K_R^* - \frac{D}{D_0} K_s^* \right) \left( \frac{D_0}{D} \right)^2 C_2 \\ & + \left[ \left( 2K_R - \frac{D}{D_0} K_s \right) \left( \frac{D_0}{D} \right)^2 C_2 \frac{\Gamma}{2} + \left( 2K_R^* - \frac{D}{D_0} K_s^* \right) \left( \frac{D_0}{D} \right)^2 \frac{C_2}{2} \right] \frac{\delta}{D} \end{aligned}$$

In the absence of salt

$$\log a_2^0 = F_2/2.3RT + \log N_2^0 - \left( K_R^* - \frac{D''}{D_0} K_s^* \right) \left( \frac{D_0}{D''} \right)^2 C_2^0 + \frac{\delta}{D} \left( 2K_R^* - \frac{D''}{D_0} K_s^* \right) \frac{C_2^0}{2}$$

in which  $D''$  is the dielectric constant of the saturated solution without salt. From the fact that in the saturated solution  $\log a_2 = \log a_2^0$  we determine the solubility as

$$\begin{aligned} \log \frac{N_2}{N_2^0} = & \left( K_R - \frac{D}{D_0} K_s \right) \left( \frac{D_0}{D} \right)^2 \frac{\Gamma}{2} + \left( K_R^* - \frac{D}{D_0} K_s^* \right) \left( \frac{D_0}{D} \right)^2 C_2 - \left( K_R^* - \frac{D''}{D_0} K_s^* \right) \left( \frac{D_0}{D_0} \right)^2 C_2^0 \\ & - \left( 2K_R - \frac{D}{D_0} K_s \right) \frac{D_0^2}{D^3} C_2 \frac{\Gamma}{2} - \left( 2K_R^* - \frac{D}{D_0} K_s^* \right) \frac{D_0^2}{2D^3} C_2^0 + \left( 2K_R^* - \frac{D''}{D_0} K_s^* \right) \frac{D_0^2}{2D^3} C_2^0 \\ = & \left[ K_R \frac{(2D' - D)}{D} - K_s \frac{D'}{D} \right] \left( \frac{D_0}{D} \right)^2 \frac{\Gamma}{2} + \left[ K_R^* \frac{D'}{D} - K_s^* \frac{(D + D')}{D} \right] \left( \frac{D_0}{D} \right)^2 C_2 \end{aligned}$$

and sodium chloride as 5.15 Å, the value of  $K_s$  calculated by the above equation is 0.14, in excellent agreement with that experimentally determined. The agreement in the case of certain other ions and dipolar ions is less satisfactory, indicating that other forces are also merged in the salting-out effect.

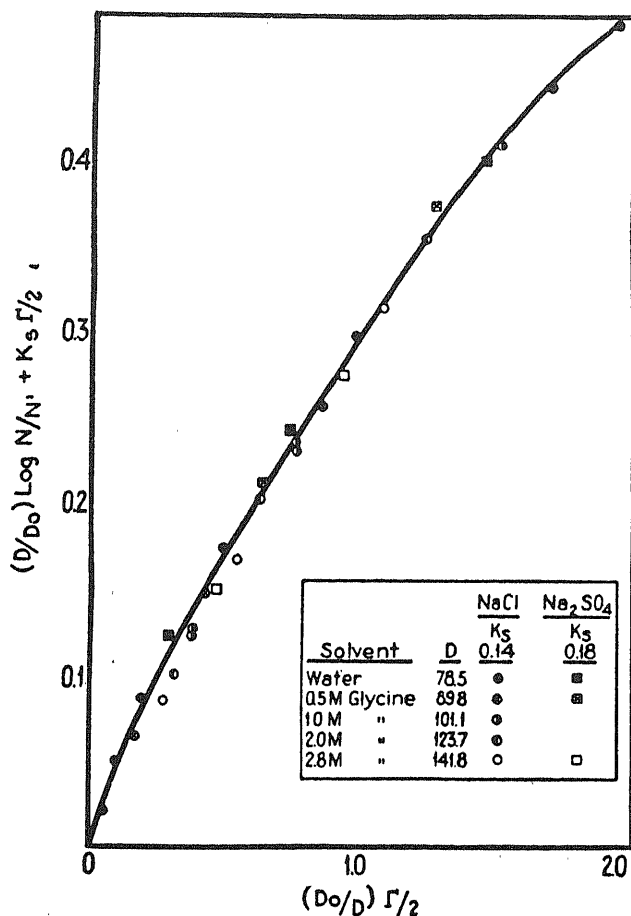


Figure 7.

Interaction between glycine, cystine and neutral salts in aqueous solutions differing in ionic strength and in dielectric constant.

The interaction between cystine and NaCl has been studied, by the same authors, in the presence of  $\alpha$ -aminobutyric acid, of the dipeptide diglycine and of urea. The results are reported in Table 3<sup>18</sup>. At low values of the ionic strength these results may also be treated as though

the only effect of the dipolar ion was through its effect on the dielectric constant, but at higher ionic strengths the various dipolar ions have quite

TABLE 3. INTERACTIONS OF CYSTINE WITH DIPOLAR IONS AND NaCl AT 25°

Ionic strength of electrolyte $\Gamma/2$	Density of solution $\rho$	Dipolar ion concentration	Solubility of Cystine		$\left(\frac{D_0 \Gamma}{D} \frac{1}{2}\right)$	Logarithm of solubility ratio $\log N/N'$	$\left(\frac{N}{D_0} \log \frac{N}{N'}\right)$	$\left(\frac{D}{D_0} \log \frac{N}{N'} + K_s \frac{\Gamma}{2}\right)$
			Moles per liter $C_s$	Mole fraction $N$				
Diglycine $\delta = 70.6$ . $K_s = 0.14$								
0.0	1.02446	0.5	0.000597	0.0000111	0.0	0.0	0.0	0.0
0.5	1.04401	0.5	0.000663	0.0000124	0.345	0.048	0.0696	0.140
1.0	1.06313	0.5	0.000699	0.0000130	0.690	0.069	0.1000	0.240
2.0	1.10023	0.5	0.000731	0.0000137	1.380	0.092	0.1333	0.413
0.0	1.07164	1.4	0.000711	0.0000140	0.0	0.0	0.0	0.0
1.0	1.10864	1.4	0.000740	0.0000147	0.443	0.021	0.0474	0.187
2.0	1.14441	1.4	0.000746	0.0000149	0.886	0.027	0.0610	0.341
$\alpha$ -Aminobutyric acid $\delta = 22.6$ . $K_s = 0.14$								
0.0	1.01082	0.5	0.000475	0.00000884	0.0	0.0	0.0	0.0
0.2	1.01887	0.5	0.000541	0.00001007	0.175	0.057	0.0652	0.093
0.5	1.03068	0.5	0.000550	0.0000102	0.437	0.063	0.0721	0.142
1.0	1.04997	0.5	0.000618	0.0000115	0.874	0.115	0.1316	0.272
2.0	1.08748	0.5	0.000695	0.0000130	1.748	0.168	0.1922	0.472
0.0	1.02435	1.0	0.000461	0.00000884	0.0	0.0	0.0	0.0
0.5	1.04391	1.0	0.000526	0.0000101	0.389	0.057	0.0734	0.143
1.0	1.08260	1.0	0.000563	0.0000108	0.777	0.086	0.1108	0.251
2.0	1.09959	1.0	0.000633	0.0000122	1.554	0.139	0.1790	0.459
0.0	1.03764	1.5	0.000442	0.00000875	0.0	0.0	0.0	0.0
0.5	1.05665	1.5	0.000497	0.00000985	0.350	0.051	0.0730	0.143
0.5	1.05660	1.5	0.000494	0.00000979	0.350	0.049	0.0702	0.140
1.0	1.07511	1.5	0.000509	0.0000101	0.699	0.062	0.0888	0.229
2.0	1.11158	1.5	0.000556	0.0000111	1.398	0.103	0.1475	0.428
Urea Concentration = 2.0, $D = 84.03$ ; Concentration = 5.0, $D = 91.07$								
0.0	1.02852	2.0	0.000622	0.0000119	0.0	0.0	0.0	
0.1	1.03247	2.0	0.000661	0.0000126	0.094	0.024	0.0257	
0.25	1.03831	2.0	0.000703	0.0000134	0.234	0.051	0.0546	
0.5	1.04801	2.0	0.000727	0.0000139	0.468	0.067	0.0717	
1.0	1.06704	2.0	0.000778	0.0000149	0.935	0.097	0.1038	
0.0	1.07385	5.0	0.000791	0.0000165	0.0	0.0	0.0	
0.25	1.08312	5.0	0.000845	0.0000176	0.216	0.018	0.0209	
1.0	1.11065	5.0	0.000952	0.0000199	0.862	0.071	0.0824	

specific effects on interactions in these three-component systems, as they must, since they interact differently both with the ions and the other dipolar ions in the system.

## Solubility of Salts in Aqueous Solutions Containing Dipolar Ions

The interactions between amino acids and salts that have been considered thus far were carried out by measuring the change in solubility of the amino acid with salt concentration. This method, although it has the great advantage of simplicity of observation and interpretation has the disadvantage that interactions can be investigated at only one amino acid concentration, namely, that of the saturated aqueous solutions. A complete understanding of the interactions between ions and dipolar ions must depend upon a knowledge of the activity coefficients of both components at every concentration of either.

**Influence of the Salt.** There are a number of other methods which can be employed in the investigation of the interaction between ions and dipolar ions. Among them are (1) measurement of the solubility of a difficultly soluble salt in the presence of a dipolar ion; in studies of this kind the concentration of the amino acid or peptide can be varied widely but the ionic strength is fixed; (2) in measurements of the freezing point<sup>19</sup>; and (3) of the electromotive force of systems containing ions and dipolar ions<sup>20</sup>, both the concentration of the ions and dipolar ions can be varied widely. Studies on systems containing glycine and sodium or thallous chloride have shown that results may be obtained with these quite different procedures which, as a first approximation, are in satisfactory agreement.

The salt solubility method, which cannot be employed in the study of interactions between glycine and sodium chloride, since sodium chloride is too soluble in aqueous solution, has been used to investigate the influence of amino acids on the solubility of such salts as thallos chloride<sup>21</sup>, thallos iodate<sup>22</sup>, and other iodates and bromates<sup>23, 24</sup>.

Direct comparison is possible of the activity coefficients of thalious chloride in systems containing glycine as estimated by both the solubility and electromotive force methods (Table 4). The ionic strengths at which all of these measurements were carried out are of necessity very low, being limited by the solubility of the thalious chloride. Whether estimated by solubility or electromotive force measurements the logarithm of the coefficient of the salt,  $\gamma_3$ , appears to be nearly linear in the amino acid concentration,  $m_2$ . The value of  $(-\log \gamma_3/m_2)$  is thus nearly constant at each ionic strength and is somewhat smaller the higher the ionic strength. The values reported from electromotive force measurements are however appreciably smaller than those from solubility measurements.

<sup>19</sup> Scatchard, G., and Prentiss, S. S., *J. Am. Chem. Soc.*, **56**, 2314 (1934).

<sup>21</sup> Joseph, N. R., *J. Biol. Chem.*, **111**, 479, 489 (1935).

<sup>21</sup> Failey, C. F., *J. Am. Chem. Soc.*, 54 576 (1932).

22. H. L. Hsieh, C. H. Hsu, and C. H. Chen, *Chem. Soc.*, 54 870 (1952).

The interactions between glycine and thallos chloride have also been studied at two temperatures by the electromotive force method at the

TABLE 4. COMPARISON OF INTERACTIONS OF DIPOLAR IONS WITH TlCl ESTIMATED RESPECTIVELY BY EMF AND SOLUBILITY MEASUREMENTS

Electrolyte	Electrolyte concentration $m_2$	Amino acid concentration $m_3$	Activity coefficient of electrolyte $-\log \gamma_2$	$-\log \gamma_3/m_2$
<i>Glycine</i>				
Solubility measurements (1). $-\log \gamma = \log m_3/m_3^0$ ( $m_3^0 = 0.01617$ )				
TlCl at 25°	0.01623	0.01	0.0017	0.170
	0.01634	0.03	0.0047	0.156
	0.01648	0.05	0.0083	0.166
	0.01678	0.10	0.0162	0.162
	0.01745	0.20	0.0331	0.166
EMF measurements (2). $-\log \gamma = -\log \gamma_3/\gamma_3^0$				
TlCl at 25°	0.005	0.10	0.011	0.110
		0.20	0.021	0.105
		0.40	0.041	0.103
		1.00	0.108	0.108
	0.015	0.10	0.009	0.090
		0.20	0.017	0.085
		0.50	0.047	0.094
		1.00	0.091	0.091
		2.00	0.179	0.090
		TlCl at 1°	0.005	0.20
0.40	0.060			0.150
0.60	0.091			0.151
1.00	0.148			0.148
<i>Alanine</i>				
TlCl at 1°	0.005	0.20	0.015	0.075
		0.40	0.031	0.078
		1.00	0.067	0.067
$\alpha$ -Aminobutyric acid				
TlCl at 1°	0.005	0.40	0.023	0.058
		0.60	0.033	0.055
<i>dl-Valine</i>				
TlCl at 1°	0.005	0.20	0.005	0.025
		0.40	0.011	0.028

(1) Failey, C. F., *J. Am. Chem. Soc.*, **54**, 576 (1932).

(2) Joseph, N. R., *J. Biol. Chem.*, **111**, 489 (1935).

same temperature (20) demonstrate that, as a first approximation, glycine has a somewhat similar influence on the activity coefficients of these two salts of the same valence type, though the limiting slope ( $-\log \gamma_3/m_2$ ) for sodium chloride, 0.14, (see equation 9) is slightly smaller. Whereas these two chlorides have very nearly the same interaction constants, the solubility of thallous iodate is increased by glycine to a far greater extent than is that of thallous chloride (Table 5). The higher solvent action upon the iodate, although related in part to the lower ionic strengths in these systems, would appear to reflect also the specific properties of the anions. La Mer and Goldman<sup>25</sup> have reported a very large effect of ethyl alcohol on the solubility of thallous iodate.

In the most recent studies of the influence of dipolar ions upon difficultly soluble salts<sup>23, 24</sup>, the anions were iodates and bromates. Cations with effects of two kinds were considered. Barium and calcium iodate and barium bromate had the higher interaction constants to be expected from their higher valence. The observed increases in solubility (Table 6), like those in Tables 4 and 5, are closely linear in the concentration, the ratio  $-\log \gamma_3/m_2$  increasing with decreasing concentration and being more than twice as great as for thallous chloride, but not for thallous iodate.

In the studies of the interactions of glycine and alanine with silver and lead iodates, however, the relations are more complex, and Keefer and Reiber have presented evidence of the formation of new compounds. This study thus reflects another kind of phenomenon, which also plays a great role in protein chemistry<sup>26</sup> and which cannot completely be accounted for in terms of electrostatic interactions.

**Influence of the Dipolar Ions.** The influence of amino acids of different types on difficultly soluble salts has been investigated both by the solubility and the EMF method. The latter results upon alanine,  $\alpha$ -aminobutyric acid and *dl*-valine are collected in Table 4, the former in Tables 5 and 6. The values of the limiting slopes are in every case smaller the longer the paraffin side chain of the amino acid.

Smaller than the effect of any of the amino acids is that of the anhydride, diketopiperazine (Table 5). Its effect upon both thallous chloride and thallous iodate has been investigated and is of the order that might have been expected from its solubility in non-polar solvents and its low electrical moment. Conversely, glycylglycine, with its higher moments, gives a far greater solvent action than  $\alpha$ -amino acids. The results with  $\beta$ -alanine are less clear.

The influences of these dipolar ions upon thallous iodates are graphically

TABLE 5. INTERACTIONS OF DIPOLAR IONS WITH  $TlCl$ ,\* AND  $TlIO_3$ \* AT 25°

Dipolar ion	Dipolar ion concentration $C_2$	Ionic strength of electrolyte $\Gamma/2 = C_2$	Activity coefficient of electrolyte $-\log \gamma_{\pm}$	$-\log \gamma_{\pm}/C_2$
<i>TlCl</i>				
Glycine.....	0.2	0.01725	0.0296	0.148
$\alpha$ -Alanine.....	0.2	0.01703	0.0240	0.120
$\alpha$ -Amino- <i>n</i> -butyric acid.....	0.1	0.01648	0.0098	0.098
	0.2	0.01688	0.0203	0.102
$\alpha$ -Aminoisobutyric acid.....	0.2	0.01679	0.0180	0.090
$\alpha$ -Amino- <i>n</i> -valeric acid.....	0.1	0.01627	0.0042	0.042
	0.2	0.01633	0.0059	0.030
Urea.....	0.2	0.01642	0.0082	0.041
Diketopiperazine.....	0.1	0.01604	-0.0019	-0.019
Glycylglycine.....	0.2	0.01835	0.0566	0.280
<i>TlIO<sub>3</sub></i>				
Glycine.....	0.025	0.001875	0.0075	0.300
	0.05	0.001907	0.0149	0.298
	0.10	0.001971	0.0292	0.292
	0.15	0.002033	0.0423	0.282
	0.20	0.002098	0.0563	0.282
$\alpha$ -Alanine.....	0.025	0.001872	0.0068	0.272
	0.05	0.001907	0.0149	0.298
	0.10	0.001963	0.0275	0.275
	0.20	0.002084	0.0534	0.267
Sarcosine.....	0.05	0.001880	0.0087	0.174
	0.10	0.001917	0.0171	0.171
	0.15	0.001954	0.0254	0.169
	0.20	0.001993	0.0341	0.170
Dimethylglycine.....	0.025	0.001859	0.0038	0.152
	0.05	0.001873	0.0070	0.140
	0.10	0.001901	0.0135	0.135
	0.20	0.001957	0.0261	0.131
$\alpha$ -Amino- <i>n</i> -butyric acid.....	0.10	0.001906	0.0146	0.146
	0.20	0.001971	0.0292	0.146
$\alpha$ -Aminoisobutyric acid.....	0.05	0.001874	0.0072	0.144
	0.10	0.001901	0.0135	0.135
	0.15	0.001934	0.0210	0.140
	0.20	0.001956	0.0258	0.129
$\alpha$ -Amino- <i>n</i> -valeric acid.....	0.10	0.001895	0.0122	0.122
	0.20	0.001939	0.0221	0.111
Urea.....	0.025	0.001845	0.0005	0.020
	0.05	0.001851	0.0019	0.038
	0.10	0.001861	0.0042	0.042
	0.15	0.001876	0.0077	0.051
	0.20	0.001891	0.0112	0.056
Diketopiperazine.....	0.10	0.001844	0.0002	0.002
$\beta$ -Alanine.....	0.025	0.001863	0.0047	0.188
	0.05	0.001884	0.0096	0.192
	0.10	0.001919	0.0175	0.175

represented in Fig. 8, amino acid concentrations and solubilities being given as moles per liter. Less readily compared with the electromotive force measurements than the results upon thallous chloride in Table 4 these studies upon interactions with thallous iodate illustrate (1) the closely linear relation between  $-\log \gamma_3$  and concentration at very low ionic strengths, as well as the profound influence of (2) non-polar groups in diminishing and (3) of the dipole moment in increasing the interactions between ions and dipolar ions.

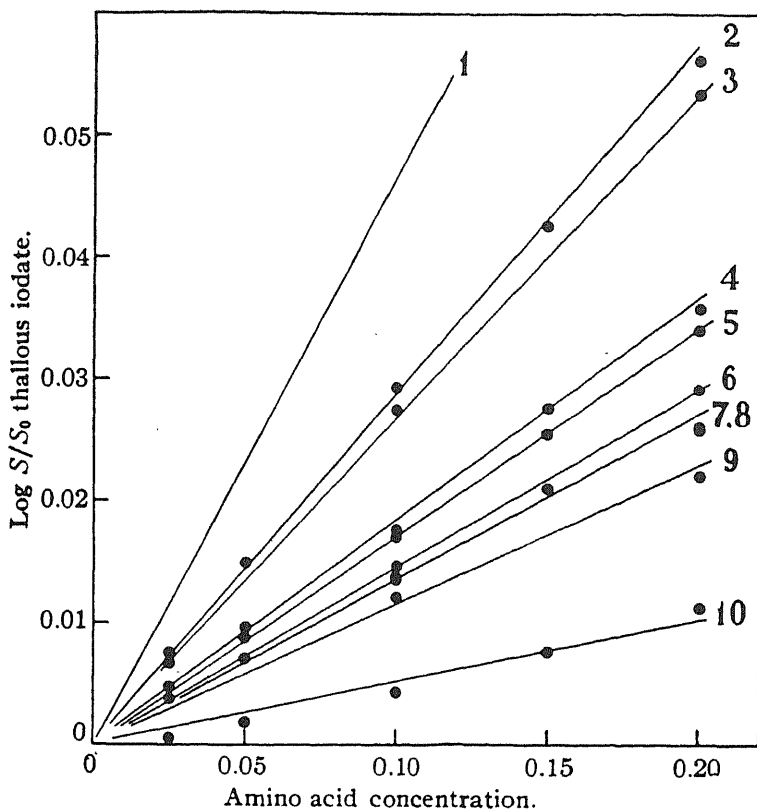


Figure 8. 1, Glycylglycine; 2, glycine; 3,  $\beta$ -alanine; 4,  $\alpha$ -alanine; 5, sarcosine; 6,  $\alpha$ -amino-*n*-butyric acid; 7, dimethylglycine; 8,  $\alpha$ -aminoisobutyric acid; 9,  $\alpha$ -amino-*n*-valeric acid; 10, urea. From C. F. Failey, *J. Am. Chem. Soc.*, 55, 4374 (1933).

#### \* Freezing Point and EMF Measurements in Aqueous Salt Solutions Containing Dipolar Ions

Freezing point and EMF measurements upon systems containing both



TABLE 6. SOLUBILITY OF AMINO ACIDS IN  $\text{Ba}(\text{IO}_3)_2$ ,  $\text{Ca}(\text{IO}_3)_2$ ,  $\text{Ba}(\text{BrO}_3)_2$ ,  $\text{AgIO}_3$  AND  $\text{Pb}(\text{IO}_3)_2$ 

Amino acid concentration $m_2$	Electrolyte concentration $m_1$	$-\log \gamma_2$	$-\log \gamma_1/m_2$
<i>Glycine in <math>\text{Ba}(\text{IO}_3)_2</math></i>			
0.0	0.000811	0.0	
0.0251	0.000831	0.0106	0.422
0.0503	0.000851	0.0209	0.416
0.0755	0.000871	0.0310	0.411
0.1008	0.000895	0.0428	0.425
0.1990	0.000977	0.0809	0.406
0.8175	0.001552	0.2819	0.345
<i>Glycine in <math>\text{Ca}(\text{IO}_3)_2</math></i>			
0.0	0.00786	0.0	
0.0251	0.00806	0.0109	0.434
0.0503	0.00823	0.0200	0.398
0.0755	0.00849	0.0335	0.444
0.1008	0.00865	0.0416	0.413
0.2009	0.00951	0.0828	0.412
0.4055	0.01111	0.1503	0.371
0.6140	0.01297	0.2176	0.354
0.8261	0.01495	0.2793	0.338
<i>Glycine in <math>\text{Ba}(\text{BrO}_3)_2</math></i>			
0.0	0.02008	0.0	
0.0251	0.02045	0.0079	0.315
0.0503	0.02081	0.0155	0.308
0.0755	0.02113	0.0221	0.293
0.1008	0.02150	0.0296	0.294
<i>Glycine in <math>\text{AgIO}_3</math></i>			
0.0	0.0001794	0.0	
0.02510	0.0001859	0.0154	0.614
0.03768	0.0001880	0.0203	0.539
0.05026	0.0001916	0.0284	0.566
0.07536	0.0001971	0.0408	0.541
0.08234	0.0001979	0.0426	0.517
0.10075	0.0002020	0.0515	0.511
0.1238	0.0002082	0.0646	0.522
0.1654	0.0002181	0.0848	0.513
0.2042	0.0002223	0.0931	0.456
<i>Glycine in <math>\text{Pb}(\text{IO}_3)_2</math> (Sample 1)*</i>			
0.0	0.0000534	0.0	
0.00502	0.0000547	0.0105	2.092
0.01003	0.0000563	0.0230	2.293
0.01505	0.0000565	0.0345	2.292
0.02510	0.0000615	0.0614	2.446
0.05026	0.0000640	0.0787	1.566
0.08052	0.0000675	0.1018	1.264
0.10074	0.0000751	0.1481	1.470
0.1210	0.0000770	0.1590	1.314
0.2023	0.0000941	0.2461	1.216
<i>(Sample 2)*</i>			
0.0	0.0000361	0.0	
0.0249	0.0000405	0.0500	2.008
0.0499	0.0000445	0.0909	1.822
0.0749	0.0000481	0.1246	1.664
0.1000	0.0000509	0.1492	1.492

TABLE 6 (Continued).—SOLUBILITY OF AMINO ACIDS IN  $\text{Ba}(\text{IO}_3)_2$ ,  $\text{Ca}(\text{IO}_3)_2$ ,  $\text{Ba}(\text{BrO}_3)_2$ ,  $\text{AgIO}_3$ , AND  $\text{Pb}(\text{IO}_3)_2$ 

Amino acid concentration $m_2$	Electrolyte concentration $m_2$	$-\log \gamma_3$	$-\log \gamma_3/m_2$
<i>Alanine in <math>\text{Ba}(\text{IO}_3)_2</math></i>			
0.0	0.000811	0.0	
0.0251	0.000829	0.0096	0.382
0.0503	0.000843	0.0168	0.334
0.0755	0.000858	0.0245	0.325
0.1008	0.000876	0.0335	0.332
<i>Alanine in <math>\text{Ca}(\text{IO}_3)_2</math></i>			
0.0	0.00786	0.0	
0.0251	0.00800	0.0077	0.307
0.0503	0.00814	0.0152	0.302
0.0755	0.00829	0.0232	0.307
0.1008	0.00845	0.0315	0.313
<i>Alanine in <math>\text{AgIO}_3</math></i>			
0.0	0.0001794	0.0	
0.02511	0.0001892	0.0231	0.920
0.03740	0.0001924	0.0303	0.810
0.05616	0.0002001	0.0473	0.842
0.07557	0.0002064	0.0608	0.805
0.10092	0.0002155	0.0795	0.788
<i>Alanine in <math>\text{Pb}(\text{IO}_3)_2</math> (Sample 1)*</i>			
0.0	0.0000534	0.0	
0.01081	0.0000556	0.0176	1.628
0.01521	0.0000565	0.0245	1.611
0.02163	0.0000569	0.0276	1.276
0.02704	0.0000608	0.0564	2.086
0.05419	0.0000630	0.0718	1.325
0.08142	0.0000717	0.1280	1.572
<i>(Sample 2)*</i>			
0.0	0.0000361	0.0	
0.0248	0.0000405	0.0500	2.016
0.0495	0.0000466	0.1109	2.240
0.075	0.0000506	0.1467	1.956
0.100	0.0000557	0.1884	1.884

chloride and these systems have also been investigated at closely the same temperature by the electromotive force method in cells without liquid junction (reference 20, Table 4). Comparison of these results can be made in a number of ways and is facilitated by the graphical representation of the data in Fig. 9.

**Decrease in Activity Coefficient of the Salt.** The results as here plotted give the effect of the glycine,  $m_2$ , on the activity coefficients of the sodium chloride,  $\gamma_3$ . The latter are related to EMF measurements by the relation

$$\text{EMF} = -\frac{\nu RT}{2.303NF} \log \gamma_3/\gamma_3^0 \quad (8)^{26a}$$

Both methods of study

chloride (increases the value of  $-\log \gamma_3$ ). This is the reciprocal of the relation noted on the basis of solubility measurements with amino acids as saturating bodies, namely, that low concentrations of sodium chloride diminish the activity coefficients of glycine and certain other dipolar ions, such as asparagine and cystine, though not of leucine, which is salted-out by even low concentrations of sodium chloride. The salting-out effect is illustrated also in Fig. 9, however, for at the higher ionic strengths and glycine concentrations there is pronounced deviation from the linear

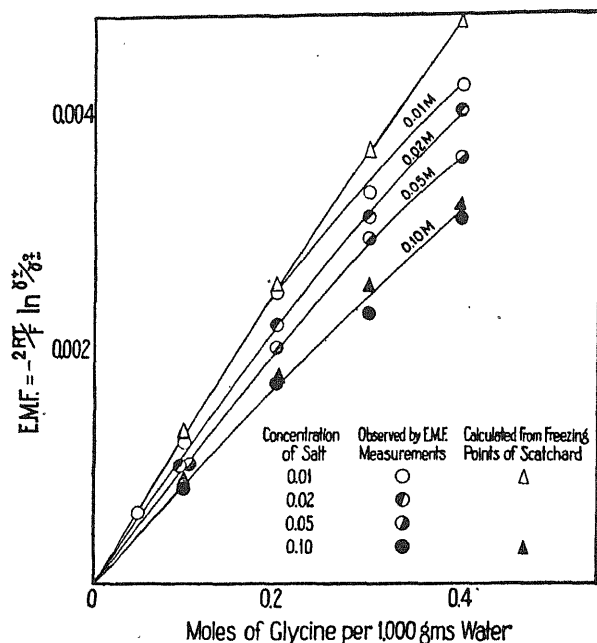


Figure 9. Influence of glycine upon dilute NaCl at 1.4°. From Joseph, N. R. *J. Biol. Chem.*, 111, 489 (1935).

relation illustrated in Fig. 8; the value of  $-\log \gamma_3$  of the salt being smaller the higher the ionic strength.

The freezing point measurements have been carried out over a wide range both of glycine and sodium chloride concentrations. Sodium chloride concentrations up to 1 molal and glycine concentrations up to 2 molal have been investigated. The results are expressed as the osmotic coefficient in the form of equation 32 of Chapter 3. This and equation 31 of that chapter give the activity coefficient of the salt as:

$$\log \gamma_3 = \frac{1}{2} \left( \frac{1}{\phi} - 1 \right) \quad \text{for } \phi = 0.9710, \quad 0.97100, \quad 3/2, \quad 0.9699, \quad 2$$

This equation computed for zero, half and molal sodium chloride yields the activity coefficients of sodium chloride in the presence of glycine, in the following relations:

$$\text{When } m_3 = 0, \quad \log \gamma_3/\gamma_3^0 = -0.14173m_2 + 0.03234m_2^2 \quad (10)$$

$$m_3 = 0.5, \quad \log \gamma_3/\gamma_3^0 = -0.04764m_2 - 0.00002m_2^2 \quad (11)$$

$$m_3 = 1.0, \quad \log \gamma_3/\gamma_3^0 = -0.01528m_2 - 0.01343m_2^2 \quad (12)$$

in which  $\gamma_3^0$  is the activity coefficient of the salt at the same salt concentration but in the absence of glycine. The equations are valid within the range of the freezing point measurements, that is to say, for values of  $(m_2 + 2m_3)$  smaller than 2. They show that increase of salt concentration decreases rapidly the effect of glycine on the activity coefficient of the

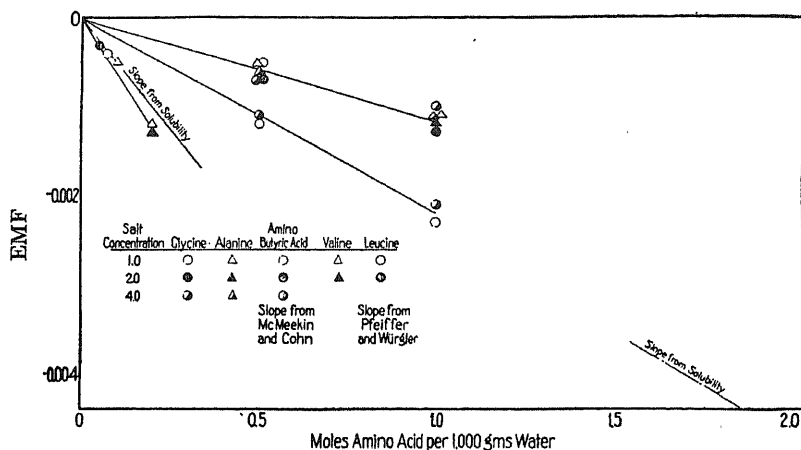


Figure 10. Influence of amino acids upon concentrated NaCl at 25°. From Joseph, N. R., *J. Biol. Chem.*, 111, 489 (1935).

salt. In dilute salt solutions, increase of the glycine concentration also decreases the effect.

**Increase in Activity Coefficient of the Salt.** The EMF measurements at 1.4° in the range of glycine concentrations to 0.5 and salt concentrations to 0.1 molal, are graphically represented in Fig. 9 and may be expressed by equation 9. EMF measurements carried out at 25°, where they could more conveniently be compared with solubility measurements, extend to concentrations of 1 molal amino acid and 4 molal salt (20). These are graphically represented in Fig. 10. Glycine, alanine, valine, α-amino-butyric acid and leucine have been studied in this way. The results indi-

for glycine and greater with increasing length of the paraffin side chain, and is related to the salting-out phenomenon.

**Increase in Activity Coefficient of Dipolar Ions.** In estimating the activity coefficients of amino acids from their effects on the activity coefficients of salts one may employ a relation derived by Bjerrum<sup>27</sup> according to which "in any system the activity coefficients of all components must satisfy equations of the form

$$\frac{\partial \ln \gamma_i}{\partial m_k} = \frac{\partial \ln \gamma_k}{\partial m_i} \quad (13)$$

"This equation follows from the definition of activity coefficients and from the identities

$$\frac{\partial^2 F}{\partial m_i \partial m_k} = \frac{\partial \bar{F}_i}{\partial m_k} = \frac{\partial \bar{F}_k}{\partial m_i} \quad (14)$$

where  $F$  is the free energy, and  $\bar{F}_i$  and  $\bar{F}_k$  are the partial molal free energies of the components  $j$  and  $k$ .

TABLE 7. VALUES OF  $(\partial \log \gamma_2 / \partial m_3)$  AT HIGH CONCENTRATION OF SODIUM CHLORIDE\*

Dipolar ion	Calculated from	
	EMF measurements	Solubility measurements
Glycine.....	0.02	0.00 (1)
Alanine.....	0.02	
$\alpha$ -Aminobutyric acid.....	0.035	0.04 (2)
Valine.....	0.09	
Leucine.....	0.12	0.09 (1)

\* From Joseph, N. R., *J. Biol. Chem.*, 111, 489 (1935).

(1) Pfeiffer, P., and Angern, O., *Z. physiol. Chem.*, 135, 16 (1924).

(2) Cohn, E. J., McMeekin, T. L., and Weare, J. H., unpublished data.

"For the case in which the component  $j$  is a strong electrolyte dissociating into  $\nu$  ions, equation 13 becomes

$$\nu \frac{\partial \ln \gamma_i}{\partial m_k} = \frac{\partial \ln \gamma_k}{\partial m_i} \quad (15)$$

where  $\gamma_i$  is the mean activity coefficient of the ions.

"For the case of three-component systems of the type studied in this paper, we may state this equation in the form

$$\log \frac{\gamma_2}{\gamma_2^0} = \int_0^{m_3} \nu \frac{\partial \log \gamma_3}{\partial m_2} \partial m_3 \quad (16)$$

"Here  $\gamma_2$  denotes the activity coefficient of amino acid in the mixed solution and  $\gamma_2^0$  is that of the pure component."

and  $m_3$ ,  $\log \gamma_2/\gamma_2^0$  can be evaluated<sup>20, pp. 481-482</sup>. Estimates of the salting-out of the amino acid by sodium chloride ( $-\log \gamma_2/m_3$ ) derived from EMF measurements are compared in Table 7 with estimates from solubility measurements. As a first approximation these results may be considered quite satisfactory and may be compared with the values in Table 8 of the salting-out constant,  $K_s$ , calculated by means of equation 7.

TABLE 8. CALCULATIONS BASED ON THE RADII OF ALKALI HALIDES AND DIPOLAR IONS

Alkali halide	Dipolar ions	Radii of dipolar ions estimated as spheres $b_d$	Radii of ions estimated as spheres $b_i$	Sum of radii of ions and dipolar ions $a$	$0.125/a$	$K_s$
		Å	Å	Å		
LiCl	Glycine	2.82	1.082	3.90	0.0321	0.064
NaCl	Glycine	2.82	1.231	4.05	0.0309	
KCl	Glycine	2.82	1.381	4.20	0.0298	
LiBr	Glycine	2.82	1.138	3.96	0.0316	0.064
NaBr	Glycine	2.82	1.288	4.11	0.0304	
KBr	Glycine	2.82	1.438	4.26	0.0293	
LiI	Glycine	2.82	1.221	4.04	0.0309	0.064
NaI	Glycine	2.82	1.370	4.19	0.0298	
KI	Glycine	2.82	1.520	4.34	0.0288	
NaCl	$\alpha$ -Alanine	3.08	1.231	4.31	0.0290	0.078
NaCl	$\alpha$ -Aminobutyric acid	3.29	1.231	4.52	0.0277	0.091
NaCl	$\alpha$ -Aminovaleric acid	3.46	1.231	4.69	0.0267	0.100
NaCl	$\alpha$ -Aminocaproic acid	3.63	1.231	4.86	0.0257	0.110

**Decrease in Activity Coefficient of Dipolar Ions.** Solubility measurements revealed that glycine was dissolved, but leucine salted-out, in dilute sodium chloride solutions (Fig. 1). The salting-out constant for glycine in dilute sodium chloride could therefore only be calculated by difference from the expected salting-in in dilute sodium chloride. The change in electromotive force and freezing point measurements in Fig. 9 indicate salting-in of glycine by sodium chloride. The Bjerrum relation or the use of equations 27 and 26 of Chapter 3 such as yielded equation 9, may be employed to calculate from these measurements the activity coefficient of glycine in sodium chloride.<sup>28, 29</sup>

**Sodium Chloride and Glycine.** The same measurements that give equations 9 to 12 yield for the activity coefficients of the glycine in the presence of sodium chloride:

$$\log \gamma_2 = (-0.28346m_3 + 0.32363m_3^{3/2} - 0.24875m_3^2 + 0.10599m_3^{5/2}) \\ + (-0.09911 + 0.12935m_3 - 0.12205m_3^{3/2})m_2 + 0.01584m_2^2 \quad (17)$$

If we compute the activity coefficients of the glycine at fixed glycine concentrations we have relations for glycine activity coefficients in the presence of sodium chloride comparable to those given above for sodium chloride in the presence of glycine.

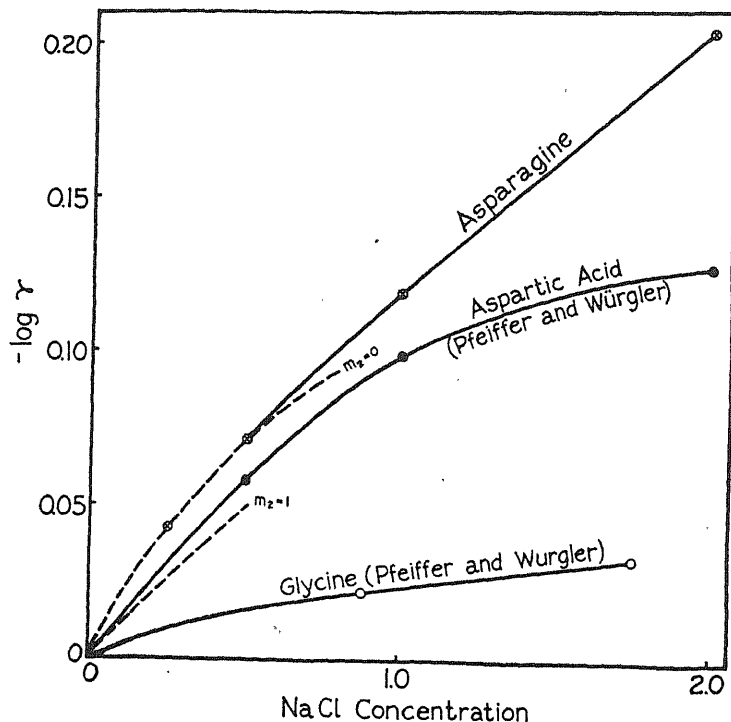


Figure 11a. Activity coefficients of amino acids in aqueous sodium chloride solutions. The dotted lines are calculated for 0 and 1M glycine concentrations from equations 18 and 19 developed by Scatchard.

When

$$m_2 = 0,$$

$$\log \gamma_2/\gamma_2^0 = -0.28346m_3 + 0.32363m_3^{3/2} - 0.24875m_3^2 + 0.10599m_3^{5/2} \quad (18)$$

$$m_2 = 1,$$

$$\log \gamma_2/\gamma_2^0 = -0.15411m_3 + 0.20158m_3^{3/2} - 0.24875m_3^2 + 0.10599m_3^{5/2} \quad (19)$$

$$m_2 = 2,$$

$$\log \gamma_2/\gamma_2^0 = -0.02476m_3 + 0.07053m_3^{3/2} - 0.24875m_3^2 + 0.10599m_3^{5/2} \quad (20)$$

sodium chloride is decreased. However measured, the interaction of ions and dipolar ions is decreased by increasing the concentration of either.

These relations in terms of the activity coefficients of the dipolar ion are graphically represented in Fig. 11a, which may be compared on the one hand with Fig. 9, on the other with Fig. 11b, representing the influ-

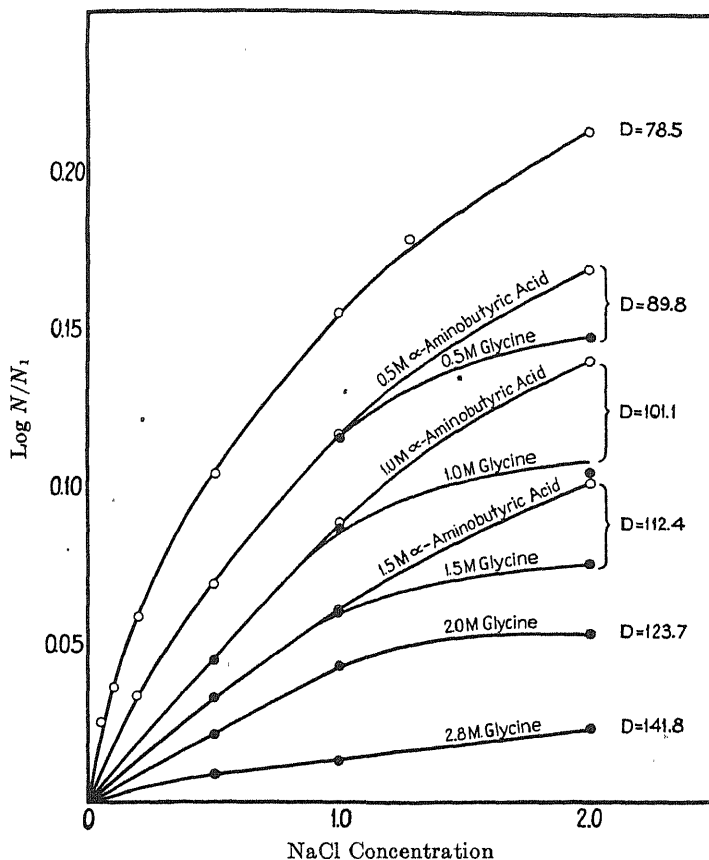


Fig. 11b. Influence of amino acids on interaction between sodium chloride and cystine.

ence of glycine on the interaction of another dipolar ion, cystine, and sodium chloride. Whether the increased concentration of the glycine is influencing its own interaction with sodium chloride or that of cystine the effect is that increase of concentration of dipolar ions reduces the interactions



measurements of less soluble amino acids, such as asparagine and cystine. The measurements of Scatchard and Prentiss (19) estimated for low glycine concentrations may, however, be so compared and values extrapolated for zero glycine ( $m_2 = 0$ ) are given in Table 9 and are graphically represented in Fig. 11a.

**Potassium Chloride and Glycine.** Measurements of the electromotive force of concentration cells with transference, together with accurate transference data, have been used by Roberts and Kirkwood<sup>30</sup> to estimate the activity of glycine in aqueous potassium chloride solutions.

TABLE 9. ACTIVITY COEFFICIENTS OF GLYCINE ( $m_2$ ) IN SODIUM CHLORIDE ( $m_3$ )

Sodium chloride		$-\log \gamma_2/\gamma_2^0$ $m_2=0$	$-\log \gamma_2/\gamma_2^0 +$ $0.064 \Gamma/2$	$-\log \gamma_2/\gamma_2^0$ $m_2$
Moles per liter $C_1 = \Gamma/2$	Moles per 1000 gm. $m_3$			
0.0499	0.05	0.0111	0.0143	0.222
0.0997	0.1	0.0206	0.0270	0.206
0.1991	0.2	0.0358	0.0485	0.179
0.3967	0.4	0.0606	0.0860	0.152
0.5930	0.6	0.0797	0.1177	0.133
0.7879	0.8	0.0937	0.1441	0.117
0.9816	1.0	0.1026	0.2008	0.103

TABLE 10. VALUES OF  $(-\partial \log \gamma_2/\partial I/2)$  IN DILUTE SOLUTION

	Molality of electrolyte	Leucine	Valine	$\alpha$ -Amino-butyric acid	Alanine	Glycine
NaCl.....	0.01	-0.09 <sup>a</sup>				0.24 <sup>b, c</sup>
TiCl <sup>b</sup> .....	0.005		0.06	0.11	0.15	0.22 <sup>d</sup> 0.30
CaCl <sub>2</sub> .....	0.01	0.04 <sup>a</sup>				0.30 <sup>e</sup>
ZnCl <sub>2</sub> <sup>b</sup> .....	0.01	0.14 0.14	0.14 0.14	0.15 0.15	0.25 0.26	0.32 0.33

From Joseph [J. Biol. Chem., 111, 479 and 489 (1935)]. Figures which are italicized refer to a temperature of  $0^\circ \pm 0.2^\circ$ ; other figures to a temperature of  $25^\circ$ .

<sup>a</sup> Amino acid solubility. J. J. Z. physiol. Chem., 97, 128 (1916)].

<sup>b</sup> EMF measurements.

<sup>c</sup> Freezing point measurements, Scatchard, G., and Prentiss, S. S., J. Am. Chem. Soc., 56, 1486 and 2314 (1934).

<sup>d</sup> Failey's [J. Am. Chem. Soc., 55, 4374 (1933)] value from TiCl solubility is 0.32 for saturated TiCl at  $25^\circ$ ; this agrees with unpublished results of Stearns, Cohn and Cohn.

Angern [Z. physiol. Chem., 135, 16 (1924)] (Joseph). It is significant only to about 0.01 of the same valence type.

For low glycine concentrations they give the activity coefficients of the glycine at ionic strengths varying from 0.05 to 0.50 by the relation

$$\log \gamma_2 = -0.1789m_3 - 0.06278m_3^2 + 0.1635m_3^{3/2} \quad (21)$$

and the limiting slope  $-\log \gamma_2/m_3$  as  $-0.18$  or appreciably lower than the limiting slopes for the interaction of glycine and sodium chloride in dilute aqueous solution. This comparison may be made with the values in

glycine and valine have also been studied by the electromotive force method<sup>20</sup>. The results graphically represented in Fig. 12 represent the influence of varying concentrations of these amino acids upon varying con-

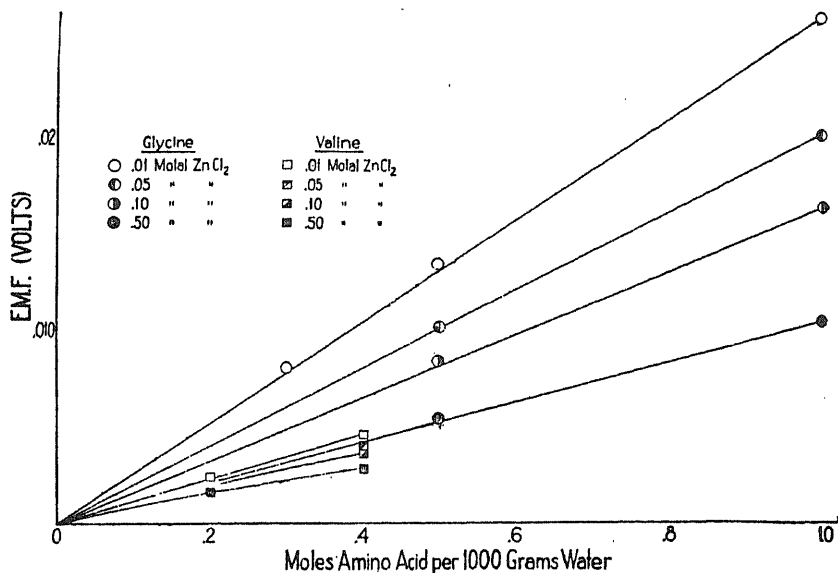


Figure 12. Influence of glycine and valine on  $\text{ZnCl}_2$  at  $1^\circ$ . From Joseph, N. R., *J. Biol. Chem.*, 111, 479 (1935).

TABLE 11. ACTIVITY COEFFICIENTS OF GLYCINE AND VALINE IN ZINC CHLORIDE SOLUTIONS AT  $1^\circ$  CALCULATED FROM EMF MEASUREMENTS<sup>1</sup>

Amino acid	Amino acid concentration $m_2$	Ionic strength of electrolyte $I/2 = 3m_2$	$(-\log \gamma_2/\gamma_2^0)$	$(-\log \gamma_2/\gamma_2^0)$	$(-\log \gamma_2/\gamma_2^0)$
Glycine	0.5	0.03	0.322	0.011	0.367
	1.0	0.03	0.322	0.011	0.367
	0.5	0.3	0.202	0.077	0.257
	1.0	0.3	0.201	0.077	0.257
	0.5	1.5	0.130	0.246	0.164
	1.0	1.5	0.129	0.246	0.164
Valine	0.2	0.03	0.140	0.004	0.133
	0.4	0.03	0.138	0.004	0.133
	0.2	0.3	0.115	0.038	0.127
	0.4	0.3	0.113	0.038	0.127
	0.2	1.5	0.085	0.153	0.102
	0.4	1.5	0.085	0.153	0.102

<sup>1</sup> Joseph, N. R., *J. Biol. Chem.*, 111, 479 (1935).

centrations of the zinc chloride. Throughout the range investigated the

## Interactions between Ions and Dipolar Ions in Alcohol-Water Mixtures

**Glycine and Sodium Chloride.** Assuming decrease in activity coefficients with increase in ionic strength to depend upon Coulomb forces we should expect these effects to become greater the lower the dielectric constant. Moreover, according to equations of the type thus far deduced, the salting-out effect<sup>4, 5</sup> should become less and less important the lower

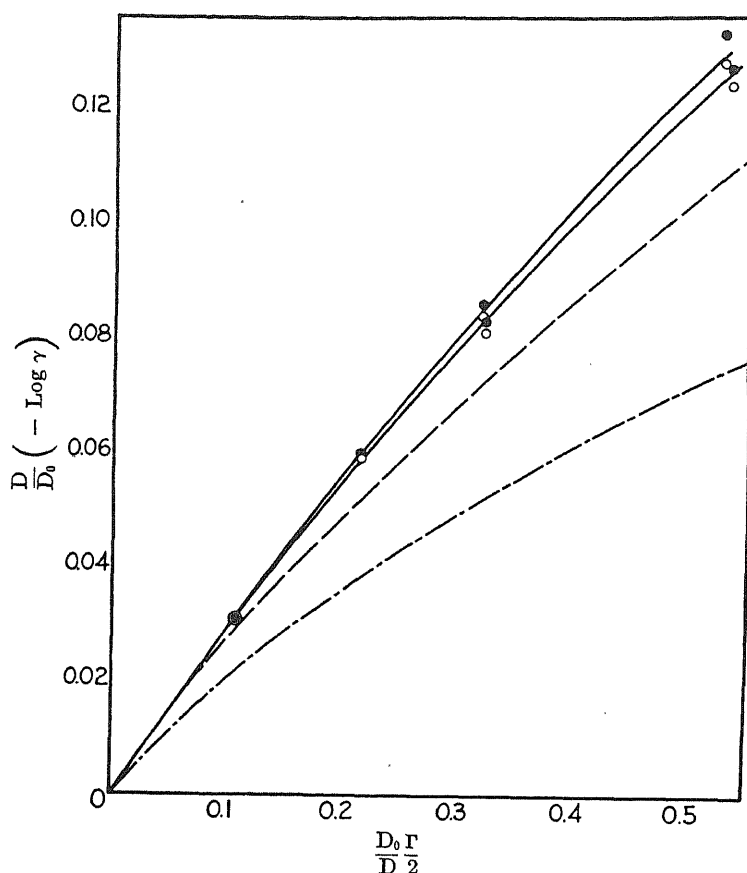


Figure 13. Activity coefficients of glycine in aqueous sodium chloride.  
 O =  $(D'/D_0) \log N/N'$ ; ● =  $(D/D_0) \log N/N'$   
 --- = Calculated from Scatchard (Table 9); - · - · - = Scatchard and Prentiss.

the dielectric constant. Studies upon the solubility of glycine in ethanol-water mixtures indicate that the salting-out effect does become less important and the influence of the dipole moment more important the larger

ments in 80% ethanol<sup>31</sup>. Since the former were obtained by extrapolation to zero glycine concentration, the dielectric constant of the solution,  $D$ , must be considered equal to that of the standard state,  $D_0$ , for these measurements. In 80% ethanol the ratio  $D_0/D$  is over 2, and but little greater if we consider the dielectric constant to be that of the pure solvent rather than of the solution, which in all but one of these measurements was less than twentieth molal with respect to glycine. The quantity  $(D_0/D)(\Gamma/2)$  is plotted in Fig. 13 against  $D/D_0$  ( $-\log \gamma$ ) both with  $D$

TABLE 12. SOLUBILITY OF GLYCINE IN 80% ETHANOL SOLUTIONS CONTAINING VARIOUS ALKALI HALIDES

Ionic strength of solvent $\Gamma/2$	Density of solution $\rho$	Ionic strength of solution $\Gamma/2$	Solubility		$\left(\frac{D_0}{D}\right)^{\frac{1}{2}}$	$\left(\frac{D_0}{D}\right)^{\frac{1}{2}}$	Log of solubility ratio $\log N/N'$	$\left(\frac{D'}{D_0} \log \frac{N}{N'}\right)$	$\left(\frac{D}{D_0} \log \frac{N}{N'}\right)$
			Moles per liter $C$	Mole fraction $N$					
Glycine in Salt-free 80% Ethanol Solution									
0.0	0.85611	0.0	0.0267	0.00102	0.0	0.0	0.0	0.0	0.0
Glycine in NaCl Solution									
0.050	0.85824	0.0498	0.0313	0.00119	0.110	0.107	0.067	0.031	0.031
0.100	0.86033	0.099	0.0361	0.00138	0.219	0.214	0.129	0.059	0.060
0.149	0.86250	0.149	0.0410	0.00156	0.328	0.319	0.184	0.084	0.086
		0.15	0.0400	0.00153	0.330	0.322	0.177	0.081	0.083
0.248	0.86664	0.247	0.0514	0.00195	0.544	0.527	0.282	0.128	0.133
		0.25	0.0497	0.00190	0.550	0.533	0.272	0.124	0.127
Glycine in KCl Solution									
0.050	0.85873	0.0498	0.0314	0.00120	0.110	0.107	0.068	0.031	0.032
0.150	0.86380	0.1492	0.0408	0.00156	0.328	0.320	0.183	0.083	0.086
Glycine in LiCl Solution									
0.100	0.85944	0.1002	0.0344	0.00131	0.220	0.216	0.108	0.049	0.050
0.249	0.86407	0.2487	0.0449	0.00170	0.547	0.532	0.221	0.101	0.103
Glycine in KI Solution									
0.100	0.86898	0.0997	0.0347	0.00132	0.219	0.215	0.112	0.051	0.052
0.250	0.88822	0.2492	0.0460	0.00176	0.548	0.533	0.235	0.107	0.110

representing the dielectric constant of the solution and of the pure solvent,  $D'$ . On the latter assumption the estimated activity coefficients due to Coulomb forces are slightly greater, but on both assumptions they are very much greater than in aqueous solution, even though this method of graphical representation should take into account those changes in Coulomb forces with change in the dielectric constant to be expected from Debye's electrostatic interaction theory.

tion 6, we could compute values of  $K_s$  (1) from the theory, or (2) from the data in water and in 80% ethanol provided we assumed the constants on the right-hand side of this equation to be the same in water and in ethanol-water mixtures.

Let us first compute  $K_s$  for glycine and sodium chloride from equation 7. Taking  $b$  for glycine as 2.82 Å and  $a$  for glycine in sodium chloride at 4.05 Å (reference 32, Table 4), the value of  $K_s$  calculated is 0.064. The limiting slope,  $K_R$ , given by equation 18 is -0.283, and  $(K_R - K_s)$  should on this basis be -0.347. Values of  $-\log \gamma + K_s(\Gamma/2)$  at finite ionic strengths may be computed from Table 9. The resulting sums, at the same values of  $(D_0/D)(\Gamma/2)$ , are actually smaller than the values of  $(D/D_0) \log N/N_0$  obtained for glycine in 80% ethanol containing sodium chloride. The difference between the limiting slopes for this interaction in water and 80% ethanol, approximately 0.08, is however of the same order as the value of  $K_s$  calculated from equation 7.

According to equation 7 the value of  $K_s$  should diminish with diminution of the dielectric constant. Thus the value of  $K_s$  for glycine in sodium chloride should diminish from 0.064 in water to 0.062 in 80% ethanol. Assuming the dielectric constant of the cavity within the dipolar ion to be greater than unity, and to have the value  $x$ , the last term in equation 7 becomes  $(D - x)/(2D + 1)$ , and  $K_s$  would be 0.061 in water and 0.055 in 80% ethanol for  $x = 5$ , and 0.056 in water and 0.046 in 80% ethanol for  $x = 10$ . Even assuming so high a value as 10 for the dielectric of the cavity does not therefore yield a sufficiently large effect to reconcile the observed activity coefficients of glycine in sodium chloride in water and in 80% ethanol, since the difference in limiting slopes to be accounted for is, as we have seen, 0.08, and the expected difference in  $K_s$  from equation 7 is 0.01.

If this effect is almost negligible, the calculation of  $K_s$  from the activity coefficients of glycine in sodium chloride in these two solvents, on the assumption that as a first approximation it is not sensibly affected by change in dielectric, leads to a more improbable result. Calculating values of  $-\log \gamma_2$  for glycine in aqueous sodium chloride, by means of equation 17 or 18, at ionic strengths equal to the values of  $(D_0/D)(\Gamma/2)$  of the sodium chloride in the ethanol-water mixtures, we have

$$K_s = \left( \frac{\log \gamma_{2 \text{ water}} - (D/D_0) \log \gamma_{2 \text{ ethanol-water}}}{\Gamma/2_{\text{water}} - \Gamma/2_{\text{ethanol-water}}} \right) \quad (22)$$

Taking  $D$  as the dielectric constant of the pure solvent,  $K_s$  becomes 0.18; or taking  $D$  as the dielectric constant of the solution, 0.19. On either

of sodium chloride with glycine than with the far larger cystine molecule for which  $K_s$  equals 0.14. The activity coefficient giving the interaction of sodium chloride and asparagine is, as we have seen, far closer to unity than should have been expected on the basis of the size of this dipolar ion. Included in the salting-out effect would appear to be forces which reflect not only the volume of the dipolar ion but also forces between its polar and non-polar groups and the surrounding ions and solvent dipoles. The simplest assumption that can be made is that at sufficiently high concentration of ethanol in the solvent the salting-out constant becomes negligible<sup>32, 33, 34</sup> and the activity coefficients approximate more closely the expectation from Coulomb's forces.

**Glycine and other Alkali Halides.** The solvent action of various alkali halides upon glycine, as well as upon most other amino acids, is greater for lithium chloride than for sodium chloride, and for sodium chloride than for potassium chloride. These observations are in accord with the theory that the limiting slope,  $-\log \gamma/C$ , as well as the curvature, should fall off with increase in the value of  $a$ ; the sum of the radii of the ions and the dipolar ions. Values for the sum of the radii of glycine and alkali halides computed from the radii of salt studied by Pauling<sup>35</sup> are listed in Table 8.

Potassium chloride is far less soluble in ethanol-water mixtures than sodium chloride, and lithium chloride far more soluble, and these differences in solubility presumably reflect interactions between ethanol and the alkali halides. If these are taken into account it is perhaps not surprising that the interactions between amino acids and alkali halides in ethanol-water mixtures are not in the same order as in aqueous solutions. The solubility ratios of glycine in 80% ethanol containing sodium chloride and potassium chloride are the same within the limits of error of present measurements. Lithium chloride, which has a larger solvent action in water has a smaller solvent action in 80% ethanol than either of these salts, and this observation<sup>32</sup> suggests that forces other than those that obtain in aqueous solutions must be taken into account in these mixed solvents (Fig. 14).

On the basis of the sum of the radii listed in Table 8 the interaction of glycine with potassium iodide should be the smallest among these alkali halides. The measurements upon the interaction of glycine and potassium iodide that have been carried out are in agreement with this expectation, the solubilities being smaller at a given concentration of this salt than at the same concentration of lithium chloride or of potassium, or sodium chloride.

**Glycine and Lithium Chloride.** Not only is glycine more soluble in aqueous solutions of lithium chloride than in the other alkali halides that have thus far been studied, but lithium chloride is more soluble in ethanol-water mixtures. The interactions between glycine and lithium chloride have been studied both in water and in ethanol-water mixtures containing as much as 95% alcohol. The solubility of glycine, in the neighborhood of 3.0 moles per liter in aqueous solution, does not exceed 0.005 mole per liter in 95% ethanol containing the largest amount of lithium chloride studied. Whereas the dielectric constants of its aqueous solutions thus greatly exceed that of pure water, the dielectric constant of solvent and solution can be considered identical in 95% ethanol, and the solubility ratios in this solvent can be considered activity coefficient ratios.

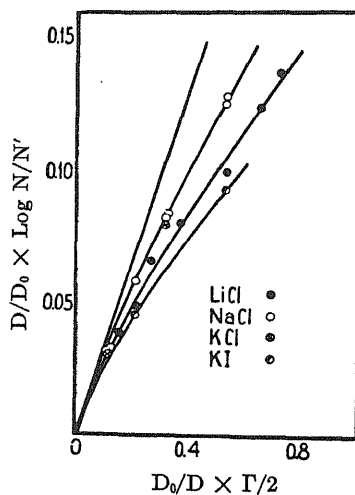


Figure 14. Interaction of glycine and alkali halides in ethanol-water mixtures.

When the measurements upon glycine in ethanol-water mixtures varying from 60 to 95%, containing lithium chloride, given in Table 13 are plotted in the same manner as those in 80% ethanol containing sodium chloride, (Fig. 15), the dielectric constant being taken as that of the pure solvent,  $D'$ , the interactions between ion and dipolar ion, as judged by the value of  $(D'/D_0) \log N/N'$  at the same value of  $(D_0/D')(\Gamma/2)$ , would appear to be greater the lower the dielectric constant. If, however, the dielectric constant is taken as that of the solution,  $D$ , then since  $D$  is greater than  $D'$  in the solutions containing more water and in which the glycine is more soluble, all the measurements from 60 to 95% ethanol-water mixtures

17), indicate that this is a fair first approximation. The results in Table 13 are represented graphically in this way in Fig. 15. Estimates of the interaction between glycine and lithium chloride that have previously been published have also been based on this method of calculation<sup>32, 33, 34</sup>.

**$\alpha$ -Aminocaproic Acid and Lithium Chloride.** Norleucine, or  $\alpha$ -aminocaproic acid, is far less soluble in water than glycine. The dielectric con-

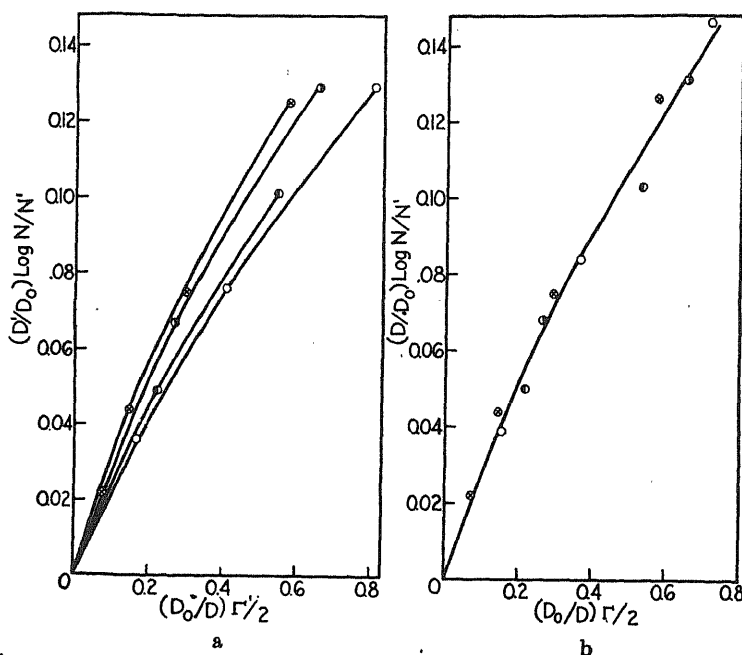
TABLE 13. SOLUBILITY OF GLYCINE IN WATER AND ETHANOL-WATER MIXTURES CONTAINING LiCl AT 25°

Ionic strength of solvent $I'/2$	Density of solution $\rho$	Ionic strength of solution $I'/2$	Solubility		$\left(\frac{L}{S}\right)_{I=0}$	$\left(\frac{L}{S}\right)_{I=0}$	Log of solubility ratio $\log N/N'$	$\left(\frac{N}{D_0} \log \frac{N}{N'}\right)$	$\left(\frac{N}{D_0} \log \frac{N}{N'}\right)$
			Moles per liter $C$	Mole fraction $N$					
Glycine in Water									
0.0	1.08283	0.0	2.8715	0.0563	0.0	0.0	0.0	0.0	0.0
0.25	1.08977	0.220	2.9626	0.0584	0.220	0.119	0.015	0.015	0.028
0.746	1.10169	0.655	3.0322	0.0600	0.655	0.350	0.027	0.027	0.051
Glycine in 60% Ethanol									
0.0	0.91074	0.0	0.1601	0.00468	0.0	0.0	0.0	0.0	0.0
0.1019	0.91418	0.101	0.1836	0.00536	0.166	0.153	0.059	0.036	0.039
0.2483	0.91895	0.2458	0.2139	0.00623	0.405	0.368	0.129	0.076	0.084
0.4951	0.92681	0.4893	0.2624	0.00764	0.806	0.716	0.213	0.129	0.146
Glycine in 80% Ethanol									
0.0	0.85611	0.0	0.0267	0.00102	0.0	0.0	0.0	0.0	0.0
0.1003	0.85944	0.1002	0.0344	0.00131	0.220	0.216	0.108	0.049	0.050
0.2492	0.86407	0.2487	0.0449	0.00170	0.547	0.532	0.221	0.101	0.103
Glycine in 90% Ethanol									
0.0	0.82521	0.0	0.00557	0.000254	0.0	0.0	0.0	0.0	0.0
0.1015	0.82879	0.1015	0.00840	0.000382	0.266	0.265	0.177	0.067	0.068
0.2511	0.83376	0.2510	0.01227	0.000555	0.658	0.652	0.339	0.129	0.131
Glycine in 95% Ethanol									
0.0	0.80712	0.0	0.00183	0.000094	0.0	0.0	0.0	0.0	0.0
0.0253	0.80823	0.0253	0.00213	0.000109	0.073	0.073	0.064	0.022	0.022
0.0504	0.80922	0.0504	0.00248	0.000126	0.146	0.146	0.128	0.044	0.044
0.1018	0.81111	0.1018	0.00306	0.000155	0.295	0.294	0.217	0.075	0.075
0.1987	0.81458	0.1987	0.00424	0.000216	0.575	0.573	0.362	0.125	0.126

stant of its solutions may be taken, as a first approximation, as close to that of the pure solvents. Moreover, whereas lithium chloride in water increases the solubility of glycine, it first increases and then diminishes the solubility of leucine; the salting-out effect being dominant even in a



parable. In 60% ethanol glycine is six times as soluble as norleucine, but in 95% ethanol their solubilities are of the same order and are very small. In these alcohol-water mixtures moreover the solvent action of lithium chloride on these two quite different amino acids is nearly the same, being only slightly smaller for the larger molecule. We must therefore conclude either that the salting-out effects, so different in water, vanish in these alcohol-water mixtures, or that the interaction of the lithium chloride with the ethanol is so large that the difference between the salting-out effects upon the two different amino acids becomes a small part of large



Figures 15a and b. Interaction of glycine and lithium chloride in ethanol-water mixtures at 25°. ○ 60% ethanol; ● 80% ethanol; ◐ 90% ethanol; ⊕ 95% ethanol.

apparent salting-out constants. In either case these results demonstrate that in solutions of low dielectric constant interactions between ions and dipolar ions even of different composition are closely similar if they have the same dipole moment.

**$\epsilon$ -Aminocaproic Acid and Lithium Chloride.** Glycine and norleucine ( $\alpha$ -aminocaproic acid) with the same moment, but different compositions, have, as we have seen, closely the same activity coefficient in 95% ethanol

totally different activity coefficients under these conditions. Measurements defining interactions of these isomers with ions in such solutions are reported in Table 14. *The interactions between ions and dipolar ions are greater the greater the dipole moment of the latter.*

Measurements upon  $\epsilon$ -aminocaproic acid have thus far been made at only one dielectric constant. They can therefore not be employed in estimating  $K_s$  in ethanol-water mixtures. They can readily be compared

TABLE 14. SOLUBILITY OF  $\alpha$  AND  $\epsilon$ -AMINOCAPROIC ACIDS IN WATER AND ETHANOL-WATER MIXTURES CONTAINING LiCl AT 25°

Ionic strength of solvent $\Gamma'/2$	Density of solution $\rho$	* Ionic strength of solution $\Gamma'/2$	Solubility		$\left(\frac{L}{N}\right)_{\infty}$	$\left(\frac{L}{N}\right)_{\infty}$	Log of solubility ratio $\log N/N'$	$\left(\frac{D'}{D_0} \log \frac{N}{N'}\right)$	$\left(\frac{D}{D_0} \log \frac{N}{N'}\right)$
			Moles per liter $C$	Mole fraction $N$					
<i>dl</i> - $\alpha$ -Aminocaproic Acid in Water									
0.0	0.99922	0.0	0.0839	0.00152	0.0	0.0	0.0	0.0	0.0
0.10	1.00157	0.099	0.0872	0.00159	0.099	0.097	0.020	0.020	0.020
0.25	1.00523	0.248	0.0881	0.00160	0.248	0.241	0.022	0.022	0.023
0.50	1.01103	0.495	0.0850	0.00155	0.495	0.484	0.009	0.009	0.009
0.746	1.01696	0.739	0.0837	0.00152	0.739	0.721	0.000	0.000	0.000
<i>dl</i> - $\alpha$ -Aminocaproic Acid in 60% Ethanol									
0.0	0.90617	0.0	0.0260	0.000757	0.0	0.0	0.0	0.0	0.0
0.100	0.90899	0.0998	0.0294	0.000856	0.164	0.162	0.053	0.032	0.033
0.250	0.91294	0.2491	0.0331	0.000963	0.410	0.404	0.105	0.063	0.065
<i>dl</i> - $\alpha$ -Aminocaproic Acid in 95% Ethanol									
0.0	0.80729	0.0	0.00254	0.000129	0.0	0.0	0.0	0.0	0.0
0.026	0.80824	0.0256	0.00303	0.000154	0.074	0.074	0.077	0.027	0.027
0.050	0.80926	0.0504	0.00340	0.000173	0.146	0.146	0.127	0.044	0.044
0.102	0.81117	0.1018	0.00414	0.000211	0.295	0.294	0.214	0.074	0.074
0.199	0.81464	0.1986	0.00553	0.000282	0.575	0.573	0.337	0.116	0.117
$\epsilon$ -Aminocaproic Acid in 95% Ethanol									
0.0	0.80800	0.0	0.0149	0.000760	0.0		0.0	0.0	
0.010	0.80843	0.00998	0.0174	0.000887	0.029		0.067	0.023	
0.026	0.80921	0.0256	0.0203	0.00103	0.074		0.130	0.046	
0.050	0.81039	0.0503	0.0252	0.00128	0.146		0.226	0.078	
0.102	0.81270	0.1014	0.0353	0.00180	0.294		0.345	0.129	

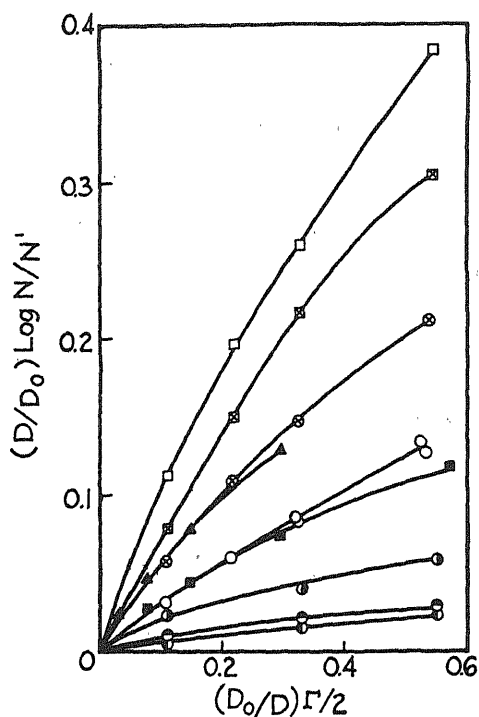
with the measurements upon glycine and  $\alpha$ -aminocaproic acid in ethanol-water mixtures by the approximate equation 5 in which the term  $K_s$  is neglected. The results for the  $\alpha$ -amino acids are fairly well represented by this equation if  $B$  is put equal to 0.95 for the interaction with lithium chloride<sup>36</sup>, in which case  $K_R$  is equal to 0.33.

**Hydantoic Acid. Pentides and Sodium Chloride.** The pentides that

have thus far been investigated both in water and ethanol-water mixtures are far less soluble than these amino acids. Their interactions with sodium chloride have been investigated in 80% ethanol<sup>31, 32</sup> and are graphically represented in Fig. 16, in which the results with glycine,  $\alpha$  and  $\epsilon$ -aminocaproic acids are included so as to facilitate comparison. *The activity coefficients both in the series of amino acids and peptides of varying dipole moment deviate further from unity the greater their dipole moments.*

Figure 16.

Solubility ratios in 80% ethanol containing sodium chloride:  $\square$  lysylglutamic acid,  $\boxtimes$  triglycine,  $\otimes$  diglycine,  $\circ$  glycine,  $\odot$  hydantoic acid of diglycine,  $\ominus$  hydantoic acid,  $\oplus$  hydantoic acid of  $\alpha$ -aminocaproic acid; and  $\blacksquare$   $\alpha$ -aminocaproic acid and  $\blacktriangle$   $\epsilon$ -aminocaproic acid in 95% ethanol containing lithium chloride.



The values deduced for the limiting slope,  $K_s$ , are not entirely independent of the methods employed in extrapolating to zero ionic strength. In the estimates that have heretofore been made the term in  $K_s$  has been neglected and the dielectric constant has been taken as that of the solution<sup>31, 32</sup>. No sensible change in these values, would follow any present change in the method of calculation, though it is to be hoped that more accurate measurements and more detailed theories will render this approach to the problem merely a first approximation.

In Table 15 the values of  $K_s$  and the dipole moments are tabulated

TABLE 15

Substance	Interaction Constant $K_R$	Dipole Moment $\mu$
Glycine.....	0.33	15
$\alpha$ -Aminocaproic acid.....	0.33	15
Diglycine.....	0.58	26
$\epsilon$ -Aminocaproic acid.....	0.7	29
Triglycine.....	0.8	32
Lysylglutamic acid.....	1.2	59

Although the values of  $K_R$  would appear, as a first approximation, to be proportional to the dipole moment of the molecules, specific chemical factors, especially the nature of their polar and non-polar groups, cannot be neglected. Measurements upon the hydantoic acids that have been studied in 80% ethanol are also graphically represented in Fig. 16. Although the interactions of ions with these molecules are small in comparison with those with dipolar ions they are not negligible. Since many of the groups of proteins are comparable to those of the hydantoic acids we must conclude that in a second approximation to the theory of interaction between ions and dipolar ions the chemical nature of the groups will have to be considered as well as the volumes and dipole moments of the molecules. These effects will presumably be larger the higher the dielectric constant of the solution and contribute greatly to the salting-out constant and also to the specific interactions in aqueous solutions which contain more than one ionic and dipolar ionic species.

## Chapter 12

### The Theoretical Interpretation of the Properties of Solutions of Dipolar Ions

BY JOHN G. KIRKWOOD

The amino acids and proteins are ampholytes which exist in solutions partly as neutral molecules and partly as positive and negative ions. The neutral molecular species predominates in isoelectric solutions of the simpler aliphatic amino acids and constitutes the major component of isoelectric solutions of most proteins. Yet the solutions of these substances exhibit properties which distinguish them sharply from ordinary non-electrolytes. It is now recognized that their characteristic properties are attributable to a unique structure, the dipolar ionic or zwitter-ionic structure proposed by Bjerrum. A dipolar ion, while bearing no net electric charge, is characterized by electric multipole moments of large magnitude. For example, the glycine dipolar ion  $\text{NH}_3^+\text{CH}_2\text{COO}^-$  should, on the basis of structural considerations, possess an electric dipole moment of about 15 Debye units, approximately ten times that of an ordinary polar molecule. Thus a dipolar ion is a superpolar molecule, surrounded by an intense electrostatic field.

As a consequence of their electrical structure, a strong electrostatic interaction between dipolar ions and real ions in solution is to be expected, which should result in anomalies in their thermodynamic properties comparable with those observed for real ions in the presence of electrolytes. The measurements of Cohn<sup>1, 2, 3, 4</sup> and his collaborators provide a large body of thermodynamic data relating to this question, which may be interpreted by means of an extension of the Debye-Hückel theory developed by Scatchard and Kirkwood<sup>5, 6</sup> to describe the electrostatic interaction of dipolar ions and real ions. The dipolar ion structure should have a marked effect on the dissociation constants of ampholytes and on acid-base equilibrium in their solutions. This effect is interpretable on the basis of an extension of Bjerrum's<sup>7</sup> theory of electrostatic interaction in the dissocia-

<sup>1</sup> Cohn, E. J., *Naturwissenschaften*, **20**, 663 (1932); *Annual Rev. Biochem.*, **4**, 93 (1935); *Chem. Rev.*, **19**, 241 (1936).

tion of polybasic acids. Finally, the large magnitude of the dipole moments of dipolar ions should be directly observable in the dielectric polarization of their solutions. This is confirmed in a striking manner by the measurements of Wyman<sup>8</sup>, Ferry<sup>9</sup>, Oncley<sup>9</sup>, Williams<sup>10</sup> and others. Their data are susceptible of quantitative interpretation on the basis of Onsager's theory<sup>11</sup> of polar dielectrics and an extension of this theory proposed by Kirkwood<sup>12</sup>. In addition to those which have been mentioned, many other properties of solutions of dipolar ions reflect their characteristic electrical structure.

In the present chapter, we shall undertake to present a systematic theoretical interpretation of those properties of dipolar ions which are characteristic of their electrical structure. On the one hand, the theory may be employed to make semi-quantitative predictions of the properties of dipolar ions of known structure, and on the other to determine structural parameters, for example, the positions of the charged groups and the dipole moment of simple dipolar ions from experimental data. In discussing the applications of the theory, we shall be chiefly concerned with the aliphatic amino acids and their peptides. Although the principles of the theory are applicable to proteins, so many parameters are required to specify the charge distribution of a protein dipolar ion that a unique determination from the measured properties of their solutions is not at present feasible. On the other hand too little is yet known about the details of protein structure to permit a probable assignment of configuration to the dipolar ionic charges.

### Thermodynamic Interaction with Electrolytes

We shall be interested in the properties of a solution containing a dipolar ion component  $i$  at a molar concentration,  $C_i$ , and  $\nu$  ionic components at molar concentrations,  $C_1 \cdots C_\nu$ , in a solvent of dielectric constant,  $D$ . The pertinent thermodynamic properties of the solution may be derived from the chemical potentials of the several components. We shall be particularly concerned with the chemical potential of the dipolar ion species, which may be written in conventional form,

$$\begin{aligned}\mu_i &= RT \log \gamma_i C_i + \mu_i^0(T, p) \\ \mu_i^0 &= \lim_{\substack{C_1 \cdots C_\nu \rightarrow 0}} [\mu_i - RT \log C_i]\end{aligned}\quad (1)$$

where  $\gamma_i$ , the activity coefficient, is defined by this equation.

<sup>8</sup> Wyman, J., Jr., and McMeekin, T. L., *J. Am. Chem. Soc.*, **55**, 908 (1933); Wyman, J., Jr., *ibid.*, **56**, 536 (1934); **58**, 1482 (1936). See Chapter 6.

We shall make the simplifying assumption that the deviation of the solution from ideal behavior is due to electrostatic intermolecular forces alone. If the solvent is idealized as a structureless dielectric continuum, the logarithm of the activity coefficient,  $\gamma_i$ , may be written as follows,

$$\log \gamma_i = C_i K_{R_{ii}} + \sum_{k=1}^{\nu} K_{R_{ik}} C_k \quad (2)$$

$$K_{R_{ik}} = \frac{N}{1000kT} \int_0^1 \int_{\omega}^v V_{ik} e^{-W_{ik}(\lambda)/kT} dv d\lambda \quad (2a)$$

where the sum extends over all  $\nu + 1$  solute components.  $V_{ik}$  is the electrostatic work required to bring the pair of molecules or ions,  $i$  and  $k$ , from infinite separation to the given configuration in the pure solvent and  $W_{ik}(\lambda)$  is the average work (potential of average force) expended in the same process in the actual solution, all charges,  $e_1 \cdots e_s$ , of molecule  $i$  having a fraction  $\lambda$  of their full values. The integration extends over all values of the relative coordinates of the pair of molecules in the volume,  $v$ , of the solution and outside a region,  $\omega$ , of molecular dimensions determined by the size and shape of the two molecules, into which intermolecular repulsion at short range prevents penetration<sup>13</sup>.

According to the Debye-Hückel theory, the potentials of average force,  $W_{ik}$ , necessary for the evaluation of the coefficients,  $K_{R_{ik}}$ , satisfy the relation,

$$W_{ik}(\lambda) = z_k e \psi_i(\lambda, \mathbf{r}_k) \quad (3)$$

where  $\psi_i$  is the average electrostatic potential at the point  $\mathbf{r}_k$  from an origin in molecule  $i$  and  $z_k e$  is the charge of ion  $k$ . The potential,  $\psi_i$ , satisfies the Poisson-Boltzmann equation of the Debye-Hückel theory, which in its linear approximation is

$$\nabla^2 \psi_i - \kappa^2 \psi_i = 0$$

$$\kappa^2 = \frac{4\pi N e^2}{1000 D k T} \sum_{k=1}^{\nu} C_k z_k^2 \quad (4)$$

outside the region of non-penetration, while in  $\omega$  Laplace's equation is satisfied:

$$\nabla^2 \psi_i = 0 \quad (5)$$

Solution of equations 4 and 5, subject to the boundary conditions of electrostatics on the surface of  $\omega$ , continuity of the potential and the normal component of the dielectric displacement, yields  $\psi$  which may be

<sup>13</sup> Equation 2, while somewhat more convenient, for our purposes, is essentially equivalent to the more usual expression, based on the Güntelberg-Müller charging process,

used in equation 2 or 2a for the calculation of the activity coefficient. Kirkwood<sup>6</sup> has employed this method to study the influence of electrolytes on the activity of spherical dipolar ions.

In this chapter we propose to use a somewhat simpler method, which, although it yields only a limiting law in which only the linear terms in the expansion of the logarithm of the activity coefficient in the concentrations of the various solute species appear, allows the treatment of dipolar ions not spherical and permits the inclusion of "salting-out" forces between an ion and a dipolar ion, arising from a repulsion between the ionic charge and an image distribution in the cavity of low dielectric constant created by the dipolar ion in the solvent. It is evident that

$$\lim_{\substack{c_1, \dots, c_i \\ \rightarrow 0}} W_{ik}(\lambda) = \lambda V_{ik} \quad (6)$$

If we are content with the linear terms in a power series in the ionic concentrations, we may therefore write equation 2 in the following form,

$$\begin{aligned} \log \gamma_i &= \sum_{k=1}^v K'_{R_{ik}} C_k \\ K'_{R_{ik}} &= \frac{N}{1000} \int_{\omega}^{\infty} (1 - e^{-V_{ik}/kT}) dv \\ K'_{R_{ik}} &= \lim_{\substack{c_1, \dots, c_i \\ \rightarrow 0}} K_{R_{ik}} \end{aligned} \quad (7)$$

For simplicity we suppose that the term in the sum, equation 2, arising from the mutual interaction of dipolar ions is negligible and thus limit ourselves to solutions in which the concentration of dipolar ions is itself small. If desired the  $K'_{R_{ii}}$  for mutual dipolar interaction may be obtained from Fuoss<sup>14</sup> calculations for dipole molecules. The development, equation 7, is only possible when the integrals defining the  $K'_{R_{ik}}$  exist. Although they diverge when the component  $i$  is a true ion, they exist when it is a dipolar ion bearing no net charge, and the expansion is suitable for our purposes. In solvents of relatively high dielectric constant at ordinary temperatures, a satisfactory approximation to the  $K'_{R_{ik}}$  may be obtained by expansion of  $1 - e^{-V_{ik}/kT}$  with retention of the first two terms alone.

$$K'_{R_{ik}} = \frac{N}{1000} \int_{\omega}^{\infty} \left[ \frac{V_{ik}}{kT} - \frac{1}{2} \left( \frac{V_{ik}}{kT} \right)^2 \right] dv \quad (8)$$

For a dipolar ion characterized by a set of charges,  $e_1 \dots e_s$ , distributed in a cavity,  $\omega_0$ , of dielectric constant,  $D_i$ , the electrical work,  $V_{ik}$ , required to bring a true ion of charge,  $z_k e$ , from infinity to a point  $\mathbf{r}_k$  from an origin fixed in the dipolar ion  $i$  is given by



where  $\underline{\psi}(r_i)$  is the electrostatic potential in the interior of  $\omega_0$  at the location of the charge,  $\epsilon_i$ , and  $\psi(r_k)$  is the potential at the point  $r_k$  exterior to  $\omega_0$ , when the ion and dipolar ion are fixed in the given configuration in the pure solvent. Self energy terms in equation 9 are to be omitted,  $V_{ik}$  vanishing at infinite separation. The potentials  $\underline{\psi}$  and  $\psi$  satisfy Laplace's equation,  $\nabla^2\psi = 0$ , both interior and exterior to the surface of  $\omega_0$ , and everywhere on this surface fulfill the usual boundary conditions:

$$\begin{aligned}\underline{\psi} &= \psi \\ D_i \mathbf{n} \cdot \nabla \underline{\psi} &= D \mathbf{n} \cdot \nabla \psi\end{aligned}$$

where  $\mathbf{n}$  is a unit vector normal to the surface. In addition, the potentials must have the singularities characteristic of the real charge distributions of the two ions. The details of the determination of the electrostatic potential for dipolar ions of several shapes are given in the appendix.

We first consider a spherical dipolar ion of radius  $b$ , the charge distribution of which may be characterized by a point dipole of moment,  $\mu$ , located at its center. The cavity,  $\omega_0$ , is thus a sphere of radius  $b$ , and the excluded region,  $\omega$ , in the integral, equation 8, is a sphere of radius  $a$ , the sum of the radii of the dipolar ion and the real ion, all real ions in the solution being spheres of the same radius. The potential energy,  $V_{ik}$ , has the following form,

$$V_{ik} = \frac{3z_k \epsilon \mu \cos \theta}{(2D + D_i)r^2} + \frac{z_k^2 \epsilon^2 b^3}{2r^4} \frac{D - D_i}{D} \sum_{n=0}^{\infty} \frac{(n+1)}{(n+2)D + (n+1)D_i} \left(\frac{b}{r}\right)^{2n} \quad (11)$$

where  $r$  is the distance between the centers of the ion and dipolar ion and  $\theta$  is the angle between the vectors,  $\mathbf{r}$  and  $\mathbf{u}$ , the dipole moment. The second member of equation 11 represents a repulsion between the real ion and an image distribution in the cavity,  $\omega_0$ , created by the dipolar ion in the solvent. A similar term, usually of much smaller magnitude, arising from the interaction of the dipolar ion and an image distribution in the ionic cavity has been suppressed.

Substitution of equation 11 in equations 7 and 8 and the neglect of small terms in  $D_i/D$  in the summation of the resulting series yields the following limiting law for the activity coefficient of the dipolar ion component of the solution,

$$\begin{aligned}\log_{10} \gamma_i &= -K_{Ri} \Gamma/2 \\ \Gamma/2 &= \frac{1}{2} \sum_{k=1}^p C_k z_k^2\end{aligned} \quad (12)$$

$$2\pi N e^2 \left( \frac{1}{2} z_1^2 + \frac{1}{2} z_2^2 + \frac{1}{2} z_3^2 + \dots \right)$$

where  $\Gamma/2$  is the ionic strength of the solution,  $\rho$  the ratio,  $b/a$ , and  $\alpha(\rho)$  a function tabulated in table 1. Insertion of numerical values for the universal constants and the introduction of the dielectric constant of water,  $D_0$ , yields at 25°:

$$K_{Ri} = K'_{Ri} - K_{Ri}^{(1)}$$

$$K'_{Ri} = 5.48 \times 10^{-3} \left( \frac{D_0}{D} \right)^2 \frac{\mu^2}{a} = 0.125 \left( \frac{D_0}{D} \right)^2 \frac{R^2}{a} \quad (13)$$

where  $R$  is the effective dipole separation,  $\mu/\epsilon$ , and  $v_i$  is equal to  $4\pi N b^3/3$ , the value at infinite dilution of the partial molal volume of the dipolar ion component under the idealization of the solvent as structureless continuum. The units employed in equation 13 are the Debye for dipole moment and the Ångstrom for the lengths  $R$  and  $a$ . The term,  $K'_{Ri}$ , gives rise to a decrease in the chemical potential of the dipolar ion by electrolyte, salting-in, while the term,  $K_{Ri}^{(1)}$ , gives rise to an increase, salting-out. The latter effect was not included in Kirkwood's previous treatment<sup>6</sup>, based on

TABLE 1

$$\alpha(\rho) = \frac{1}{3\rho^4}[(\rho^3 - 2) \log(1 + \rho) - (\rho^3 + 2) \log(1 - \rho) - 2\rho^2]$$

$\rho$	$\alpha(\rho)$
0.0	1.00
.2	1.01
.4	1.08
.6	1.21
.8	1.54
.9	1.96

the Debye-Hückel theory. If the term  $K_{Ri}^{(1)}$  is omitted in equation 12, a limiting law in exact agreement with Kirkwood's earlier one is reached.

We next consider a dipolar ion of prolate ellipsoidal shape, characterized by a charge distribution consisting of two charges,  $+\epsilon$  and  $-\epsilon$ , located at the foci. It is convenient to employ confocal elliptical coordinates in the treatment of this model. Thus a point situated at distances,  $r_1$  and  $r_2$ , from the respective foci of the ellipsoid and in a plane inclined at angle,  $\phi$ , to reference plane containing the major axis is specified by the coordinates,  $\lambda$ ,  $\eta$ ,  $\phi$ .

$$\lambda = (r_1 + r_2)/R$$

$$\eta = (r_1 - r_2)/R \quad (14)$$

where  $R$  is the interfocal distance. The cavity,  $\omega_0$ , is an ellipsoid of eccentricity equal to  $1/\lambda_0$ , and the region,  $\omega$ , into which ions cannot penetrate is for simplicity assumed identical with  $\omega_0$ , although this is strictly

ion and a real ion of charge,  $z_k e$ , is closely approximated by the following expression,

$$V_{ik} = \frac{4z_k e^2 \eta}{DR(\lambda^2 - \eta^2)} \lambda_0^{-1} \left[ \lambda_0 - \frac{(\lambda_0^2 - 1)}{2} \log \frac{\lambda_0 + 1}{\lambda_0 - 1} \right]^{-1} \quad (15)$$

Equation 15 is exact for the two limiting cases, zero and unit eccentricity and only slightly inaccurate, because of the approximate summation of an infinite series, for intermediate eccentricities. The term arising from the interaction of the ion with its image distribution in the dipolar ion cavity, included for the sphere, has been omitted in equation 15, since it introduces undue complication, in the evaluation of the integral, equation 8. Introduction of the coördinates,  $\lambda$ ,  $\eta$  and  $\phi$  as variables of integration in the latter equation leads to the following expression,

$$K'_{R,ik} = \frac{-\pi N R^3}{8000(kT)^2} \int_{\lambda_0}^{\infty} \int_{-1}^{+1} V_{ik}^2 (\lambda^2 - \eta^2) d\eta d\lambda \quad (16)$$

the term linear in  $V_{ik}$  vanishing, since it is an odd function of  $\eta$ . Equation 16, together with 7 and 15, yields for the logarithm of the activity coefficient of the dipolar ion,

$$\begin{aligned} \log_{10} \gamma_i &= -K_{R,i} \Gamma / 2 \\ K_{R,i} &= \frac{2\pi N e^4 g(\lambda_0) R}{2303(DkT)^2} \\ g(\lambda_0) &= \lambda_0^{-2} \left[ \lambda_0 - \frac{(\lambda_0^2 - 1)}{2} \log \left( \frac{\lambda_0 + 1}{\lambda_0 - 1} \right) \right]^{-1} \end{aligned} \quad (17)$$

Insertion of numerical values for the constants gives at 25°,

$$K_{R,i} = 0.167(D_0/D)^2 g(\lambda_0) R \quad (18)$$

A tabulation of  $g$  as a function of the eccentricity of the ellipsoidal cavity is given in Table 2. For constant eccentricity  $K_{R,i}$  is proportional to the first power of the distance,  $R$ , between the charges of the dipolar ion. For elongated ellipsoids near unit eccentricity,  $g$  is approximately unity. For ellipsoids nearly spherical,  $g$  may be expanded in a power series in  $R/a$ , where  $a$  is the shortest distance of either focus to the surface, equal to the radius of the sphere at zero eccentricity. The initial term of the series,  $3R/4a$ , when substituted in equation 18 yields an expression identical with  $K_{R,i}$ , equation 13, for a dipole at the center of a sphere. No counterpart of  $K_{R,i}^{(1)}$  is obtained, since we have neglected the salting-out influence of image forces in the present case. While we shall not consider the problem in detail in the present chapter, it seems reasonable to suppose the magnitude of  $K_{R,i}^{(1)}$  for an ellipsoid can be roughly estimated by

We shall also discuss a second model in which the dipolar ion is ellipsoidal in shape but characterized by a charge distribution consisting of a point dipole of moment,  $\mu$ , situated at one of the foci and parallel to the major axis. The calculations proceed in the same manner as for the first ellipsoidal model except that the potential,  $V_{ik}$ , is of the form,

$$V_{ik} = \frac{4z_k \epsilon \mu}{DR^2} \frac{(1 + \lambda \eta)}{(\lambda + \eta)^3} \lambda_0^{-1} \left[ \lambda_0 - \frac{\lambda_0^2 - 1}{2} \log \left( \frac{\lambda_0 + 1}{\lambda_0 - 1} \right) \right]^{-1} \quad (19)$$

The logarithm of the activity coefficient of the dipolar ion is finally calculated to be

$$\log_{10} \gamma_i = -K_{Ri} \Gamma / 2 \quad (20)$$

$$K_{Ri} = \frac{3\pi N \epsilon^2 u(\lambda_0)}{2303 (DkT)^2} \frac{\mu^2}{R}$$

$$u(\lambda_0) = \frac{8}{9} \frac{1 + 2(\lambda_0^2 - 1)^{-1} + 8(\lambda_0^2 - 1)^{-2}}{\lambda_0(\lambda_0^2 - 1) \left[ \lambda_0 - \frac{\lambda_0^2 - 1}{2} \log \left( \frac{\lambda_0 + 1}{\lambda_0 - 1} \right) \right]^2}$$

With numerical values for the constants at 25°,  $K_{Ri}$  is given by the following equation,

$$K_{Ri} = 5.48 \times 10^{-3} (D_0/D)^2 u(\lambda_0) \mu^2 / R \quad (21)$$

in which the Debye and the Ångstrom are the units employed. The function  $u(\lambda_0)$  is listed for several eccentricities in Table 3. In the limit of zero eccentricity of the ellipsoid  $u(\lambda_0)/R$  reduces to  $1/a$ , where  $a$  is the radius of the sphere, and equations 20 and 21 reduce to 12 and 13, except for the salting-out term.

TABLE 2

$\epsilon$	$g(\epsilon)$	$\epsilon(1 - \epsilon^2)^{-1/2} g(\epsilon)$
0.00	0.00	0.00
.20	.30	.061
.33	.49	.17
.50	.71	.39
.60	.83	.58
.70	.94	.82
.80	1.01	1.14
1.00	1.00	

We are now ready for a brief review of the interpretation of the thermodynamic interaction of dipolar ions and electrolytes by means of the equations put forward in the preceding section. For this purpose we turn to the data of Cohn and his co-workers relating to the influence of electrolytes on the solubility of the aliphatic amino acids and the peptides of glycine.

in the presence of electrolyte, the solution being assumed ideal when the dipolar ion component is present alone. We may therefore write a limiting solubility law in the following form,

$$\log_{10} (S/S_0) = K_R \Gamma/2 \quad (22)$$

where  $\Gamma/2$  is the ionic strength and  $K_R$  is given by one of equations 12, 17, or 20, according to the assumed structure of the dipolar ion. The data of Cohn confirm the form of the limiting law, in which the initial term is linear in the ionic strength, and the coefficient  $K_R$  may be obtained from the limiting slopes of his solubility curves. (See equation 4 in Chapter 11.) In order to correlate measurements in solvents of different dielectric constant, it is convenient to introduce a coefficient  $K'_R$  equal to  $(D/D_0)^2 K_R$ , related to the solubility ratio as follows,

$$K'_R = \lim_{\Gamma \rightarrow 0} \frac{d(D_0/D) \log_{10} S/S_0}{d(D/D_0) \Gamma/2} \quad (23)$$

From equations 12, 17 and 20, we remark that the theory predicts that  $K'_R$  should be independent of the dielectric constant of the solvent if salting-out

TABLE 3

$\epsilon$	$u(\epsilon)$	$\epsilon(1 - \epsilon^2)^{-1/2}/u(\epsilon)$
0.00	0.00	0.50
.20	.46	.44
.33	.99	.35
.50	3.05	.18
.60	7.43	.093
.70	22.4	.039
.80	128.1	.0089

forces due to image repulsion are of small magnitude ( $K_R^{(1)} = 0$ ). This independence is approximately confirmed by Cohn's measurements of the solubilities of the amino acids in alcohol water mixtures, from which we may conclude that the salting-out forces, as represented in the term in  $K_s$ , are of secondary importance, though not negligible.

We shall first discuss the solubility data for the simplest amino acid, glycine, on the basis of the spherical dipolar ion model. The dipole moment,  $\mu$ , may be computed from the solubility coefficient,  $K_R^{(0)}$ , by equation 13.

$$\mu = \sqrt{183 K_R^{(0)} a} \quad (24)$$

$$K_R^{(0)} = (D/D_0)^2 K_R^{(0)}$$

From the partial molal volume of glycine at infinite dilution in water and the ionic radii of Pauling and Huggins<sup>15</sup>, Cohn estimates  $a$  to be 3.90 for

stant, he determines  $K'_R$  to be 0.32. Substitution of these values in equation 24 yields a value, 15 Debye units, for the glycine dipole moment. This value is in agreement with the dipole moment, calculated on the basis of structural considerations, for the glycine dipolar ion in which the terminal  $\text{NH}_3^+$  and  $\text{COO}^-$  carry residual charges  $+\epsilon$  and  $-\epsilon$ , respectively. It is also of interest to consider the salting-out coefficient  $K_R^{(1)}$ . Its relative importance becomes greater with increasing solvent dielectric constant. Although in water the high solubility of glycine prevents the use of equation 22 without a term for the mutual dipolar ion interaction, the coefficient  $K'_R$  may be obtained from EMF measurements of Joseph<sup>16</sup> and freezing point measurements of Scatchard and Prentiss<sup>17</sup>. For glycine and sodium chloride in water at 25°,  $K'_R$  has the value 0.24. The difference between this value and  $K_R$ , roughly 0.08, gives the salting-out coefficient  $K_R^{(1)}$  on the basis of the present theory. For glycine and sodium chloride, Cohn calculates  $a$  to be 4.05 and  $\rho$  is 0.7. By linear interpolation in Table 1, we obtain for  $\alpha(\rho)$  a value 1.37. With these values and Cohn's estimate, 57 cc., for the limiting partial molal volume of glycine, corrected for solvent electrostriction, we calculate from equation 13 a value, 0.08 for  $K_R^{(1)}$ . Thus in water the salting-out contribution arising from image forces amounts to about 25% of the salting-in contribution to the solubility coefficient  $K'_R$ , arising from the interaction of the ions of the electrolyte and the true charges of the dipolar ion.

We shall now discuss glycine and its peptides on the basis of the first ellipsoidal model, in which two charges of opposite sign are situated at the foci. The calculations will be somewhat more approximate than those for glycine on the spherical model, since the ion size of the electrolyte is neglected and the salting-out influence of image forces is not included. The distance  $R$  between the foci of an ellipsoid of eccentricity  $\epsilon$  and volume  $v/N$  may be expressed in the following manner\*

$$R = p \epsilon (1 - \epsilon^2)^{-1/3} \quad (25)$$

where the length  $p$  is equal to  $(6v/N\pi)^{1/3}$  and is to be expressed in Ångströms. With numerical values for the constants, we have

$$p = 1.47v^{1/3} \quad (26)$$

where the molal volume  $v$  is in cubic centimeters. For a sphere  $p$  is the diameter. From equations 18, 25 and 26 we may write

$$\begin{aligned} K'_R &= 0.167 \epsilon g(\epsilon) (1 - \epsilon^2)^{-1/3} p \\ R &= 1.47v^{1/3} \epsilon (1 - \epsilon^2)^{-1/3} \end{aligned} \quad (27)$$

<sup>16</sup> Joseph, N. R., *J. Biol. Chem.*, **111**, 479, 489 (1935).

From the first of these equations and the experimental value of  $K'_R$  the eccentricity of the molecular ellipsoid may be obtained by linear interpolation in Table 2. From the eccentricity and the molal volume, the charge separation of the dipolar ion may then be computed from the second of the equations. Calculations for glycine, diglycine, triglycine, and  $\beta$ -alanine are summarized in Table 4. They are based on Cohn's experimental values of  $K'_R$  and on his estimates of molal volumes corrected for solvent electrostriction (2, 3). The distance, 2.8 Å, for glycine corresponds to a dipole moment of 13 which differs only slightly from the value 15 obtained on the basis of the spherical model. A significant part of the difference between the two values is due to the neglect of electrolyte ion size in the ellipsoidal calculation. It therefore appears that we cannot conclude much about the shape of the glycine dipolar ion from the influence of salts on its activity. However, because of this very insensitivity to shape, we can place considerable confidence in the value of the dipole moment computed from the salt effect. The distances 4.8 and 6.4 computed for diglycine and triglycine are considerably below the values, 6.7 and 10.2, estimated by Cohn<sup>1</sup> for an extended chain configuration. Thus

TABLE 4

	$K'_R$	$\nu$	$R$
Glycine.....	0.32	57	2.8
Diglycine.....	.58	93	4.7
Triglycine.....	.80	130	6.4
$\beta$ -Alanine.....	.43	73	3.6

our calculations suggest that the extended chain configuration is not the preferred one, but that, because of internal rotation, the average separation of the charges in these dipolar ions lies intermediate between the extended chain value and the free rotation value. Under these circumstances, the computed distances have only formal significance as average distances unless a single preferred configuration should happen to dominate all others in probability. To take internal rotation properly into account, we should compute a  $K'_R$  for each internal configuration and then average over all configurations with an appropriate distribution function, for comparison with the experimental value of  $K'_R$ . At present it is not possible to do this, and, moreover, the ellipsoidal model would be an extremely rough approximation for "crumpled" configurations.

Finally we shall discuss a series of the aliphatic alpha-amino acids on the basis of the second ellipsoidal model in which a point dipole is located at one focus. We represent glycine by a sphere with a point dipole at its center 2.8 Å from the surface of the molecule, and the dipole moment is 13.

major axis. The eccentricity of the ellipsoid is then determined by the relation

$$(1 - e)(1 - e^2)^{-1/3} = 2l/p \quad (28)$$

where  $p$  is defined by equation 26 in terms of the molecular volume, and  $l$ , the distance of a focus from the surface is  $2.8 \text{ \AA}$  in the subsequent calculations. Equation 28 may be roughly solved by linear interpolation in Table 5. From equations 21 and 25 the following relation between the dipole moment of the dipolar ion is obtained

$$\mu^2 = 183K'_R p e (1 - e^2)^{-1/3} / u(e) \quad (29)$$

Calculations based upon equation 29 and Cohn's values of  $K'_R$  and molecular volumes for glycine,  $\alpha$ -alanine,  $\alpha$ -aminobutyric acid and leucine are listed in Table 6. The discrepancy between the dipole moment values 13 and 15, both based upon the spherical model, is merely due to the

TABLE 5

$2(l/p)$	$e$	$2(l/p)$	$e$
1.00	0.00	0.46	0.60
.81	.20	.38	.70
.70	.33	.28	.80
.55	.50		

TABLE 6

	$K'_R$	$l$	Dipole moment
Glycine.....	0.32	5.6	13
$\alpha$ -Alanine.....	.31	6.1	13
$\alpha$ -Aminobutyric acid.....	.31	6.6	13
Leucine.....	.30	7.3	13

neglect of electrolyte ion size in the present calculation, a procedure which we see does not lead to great error. It is interesting that the calculated dipole moments of glycine and the other aliphatic  $\alpha$ -amino acids turn out to have identical values. This is essentially what would be expected on the basis of structural considerations, although small differences could possibly arise from induction effects in the different aliphatic chains. The problem of internal rotation does not enter explicitly into the determination of the dipole moments of the  $\alpha$ -amino acids as it did in the case of the peptides of glycine. However, the average configuration of the aliphatic chain to which the glycine residue is attached will determine how closely the actual molecule conforms to the ellipsoidal shape, at best an approximation.<sup>17a</sup>

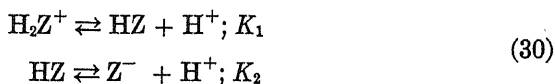
### Acid-Base Equilibrium

Important information concerning the structure of the amino acids and



ampholytes. In fact, one of the early arguments in favor of dipolar ion hypothesis was based upon the striking discrepancy between the values of the dissociation constants of the amino acids and those of aliphatic carboxylic acids. The acidic group of the neutral ampholyte was concluded to be  $\text{NH}_3^+$  rather than  $\text{COOH}$ , in agreement with a dipolar ionic structure  $\text{NH}_3^+\text{RCOO}^-$ . On the basis of Bjerrum's<sup>7</sup> theory of electrostatic interaction in the dissociation of polybasic acids, it is further to be expected that the negative group  $\text{COO}^-$  would have a marked influence on the acidic dissociation of the  $\text{NH}_3^+$  group, depending upon the distance of separation of the charged groups. From a comparison of the dissociation constants of the dipolar ion  $\text{NH}_3^+\text{RCOO}^-$  and the ion of its ester salt  $\text{NH}_3^+\text{RCOOCH}_3$ , the charge separation in the dipolar ion may be calculated. Neuberger<sup>18</sup> has calculated the charge separation for several aliphatic amino acids on the basis of the Bjerrum theory. The distances so obtained were considerably too small. More recently Westheimer and Shookhoff<sup>19</sup>, using a more refined electrostatic theory, have computed charge separations which agree very satisfactorily with structural estimates from accepted interatomic distances and bond angles. Their calculations thus complete the argument for dipolar ionic structure based upon the magnitude of dissociation constants. It also is possible to extend the electrostatic theory to an ampholyte with an arbitrary number of acidic groups,  $\text{NH}_3^+$  and  $\text{COOH}$ , and thus to provide a semi-quantitative theory of acid-base equilibria in solutions of proteins.

The ionization equilibria of a simple ampholyte  $\text{HZ}$  with positive ion  $\text{H}_2\text{Z}^+$  and negative ion  $\text{Z}^-$  are describable by the equations;



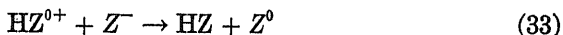
Let  $\text{HZ}^{0+}$  be a monobasic acid with a basic part,  $\text{Z}^0$ , differing in structure from the ion  $\text{Z}^-$  only by the absence of the negative charge. Its ionization equilibrium is described by



From thermodynamics we may write

$$RT \ln K_2^0/K_2 = -\Delta F^0\tag{32}$$

where  $\Delta F^0$  is the standard free energy increment in the reaction,



From a molecular point of view,  $\Delta F^0/N$  is equal to average work expended in the transport of a proton from the molecule  $\text{Z}^0$  to the ion  $\text{Z}^-$ . Since

forces bonding the protons in the acids  $\text{HZ}^{0+}$  and  $\text{HZ}$  may be expected to cancel in the transfer of the proton between them, leaving the electrostatic interaction between the proton and the negative charge of  $\text{Z}^-$  as the dominant contribution to  $\Delta F^0$ . If  $R$  is the distance between proton and the negative charge in the molecule  $\text{HZ}$ ,  $\Delta F^0$  then has the form

$$-\Delta F^0 = N\epsilon^2/D'R \quad (34)$$

where  $D'$  is an effective dielectric constant. In the Bjerrum theory  $D'$  was assumed identical with the macroscopic dielectric constant  $D$  of the solvent. In the Kirkwood-Westheimer theory<sup>20</sup> the molecule  $\text{HZ}$  is treated as a cavity of low dielectric constant rather than as a structureless system of two point charges in the solvent continuum. Using classical electrostatic theory, these authors have calculated  $D'$  for molecules of

TABLE 7

$\cos \theta$ $x$ ( $x = 4r/b$ )	-1	-1/2	0
	Values of $\sqrt{x} D'$		
0.10	0.25	0.24	0.23
0.20	0.64	0.60	0.55
0.30	1.31	1.16	1.02
0.40	2.45	2.11	1.76
0.50	4.59	3.82	3.02
0.60	8.79	7.15	5.42
0.70	17.6	14.2	10.6
0.80	37.5	30.9	23.3
0.90	79.2	68.9	56.4
1.00	127.0	116.3	103.5

spherical and ellipsoidal shape. From equations 32 and 34 the quantity  $\Delta pK$ , equal to  $\log_{10} K_2^0/K_2$ , may be computed from the formula

$$\Delta pK = \frac{\epsilon^2}{2.303 D' k T} \cdot \frac{1}{R} \quad (35)$$

where  $D'$  may be obtained from the tables of Kirkwood and Westheimer<sup>20</sup>. In their spherical model, the molecules  $\text{HZ}$  and  $\text{HZ}^{0+}$  are regarded as spheres of radius  $b$ , each containing a proton at a distance  $r$  from their centers. The molecule  $\text{HZ}$  is assumed to contain a negative charge  $e$ , in excess of  $\text{HZ}^{0+}$ , situated at a distance,  $r$ , from the center on a vector making an angle,  $\theta$ , with the vector to the proton. The values of  $D'$  in the solvent water at 25° are presented in Table 7. They were calculated with an internal molecular dielectric constant equal to 2.00. For the calculation of  $R$  from experimental values of  $\Delta pK$  it is convenient to write equation 35 in the form,

$$xD'(x) = \frac{\epsilon^2}{4.606bkT\Delta nK \sin \theta/2};$$

where  $v$  is the molar volume of the acid HZ. With an appropriate structural assignment of  $\theta$ , the function  $\sqrt{x}D'(x)$  may be computed and the corresponding value of  $x$  obtained from Table 7. From the latter value the interchange distance  $R$  is calculated from the third of equations 36.

If the acid HZ is identified with the dipolar ion  $\text{NH}_3^+\text{RCOO}^-$  and the acid  $\text{HZ}^{0+}$  with the ester salt ion  $\text{NH}_3^+\text{RCOOCH}_3$ , the theory may be used to calculate the distance between the charged groups of the dipolar ion. Such calculations have been made by Westheimer and Shookhoff, using the spherical model for glycine and alanine and the ellipsoidal model for amino acids and peptides of longer chain length. Their results are presented in Table 8 together with the distance  $R_F$  computed on the basis of free rotation and  $R_B$  on the simple Bjerrum theory in which  $D'$  is identified with the solvent dielectric constant. The distance 4.05 Å obtained for glycine

TABLE 8. CHARGE SEPARATION IN DIPOLAR IONS FROM  $\Delta pK$  VALUES

	$\Delta pK$	$R$	$R_F$	$R_B$
Glycine.....	2.02	4.05	3.56	1.53
Alanine.....	2.07	3.85	3.56	1.50
$\beta$ -Alanine.....	1.06	5.15	4.19	2.92
$\gamma$ -Aminobutyric acid.....	0.72	6.10	4.72	4.31
$\delta$ -Amino valeric acid.....	0.62	6.55	5.19	5.00
$\epsilon$ -Aminocaproic acid.....	0.38	7.85	5.63	8.16
Glycylglycine.....	0.56	6.50	5.17	5.54

TABLE 9. THE ELECTRIC MOMENT  $\bar{\mu}$  FROM DIELECTRIC CONSTANT INCREMENTS IN AQUEOUS SOLUTION

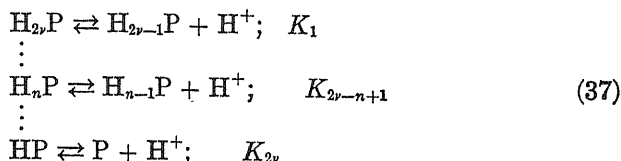
	$\delta_0$	$\mu \times 10^{18}$	$R(\text{\AA})$
Glycine.....	22.6	15.7	3.27
$\beta$ -Alanine.....	34.6	19.4	4.04
$\gamma$ -Aminobutyric acid.....	52	23.8	4.96
$\delta$ -Aminovaleric acid.....	63	26.2	5.45
$\epsilon$ -Aminocaproic acid.....	77.5	29.0	6.04
Glycylglycine.....	70.6	27.7	5.76

For reference to the values of  $\delta_0$  employed here see Chapter 6, Table 3.

agrees moderately well with the structural value 3.17. The mean charge separations in the amino acids of longer chain length, as those computed from salting-in and from the dielectric constant increment, do not differ greatly from the free rotation values. Finally, we may say that the dissociation constant data on amino acids and their ester salts are entirely in harmony with the dipolar ion hypothesis.

It is possible to extend the ideas which have just been discussed to ampholytes containing an arbitrary number of acidic and basic groups, by a method similar to that proposed by Linderstrøm-Lang<sup>41</sup>. We shall begin with some preliminary remarks on the thermodynamic aspects of the dissociation of an ampholyte. If  $D$  is a sample of formic acid in a solution

$H_{2\nu}P \dots H_nP \dots P$ , with respective charges  $n - \nu$ . The dissociation equilibria are described by the equations:



If  $(H^+)$  is the hydrogen ion activity,  $C^0$  the bulk ampholyte concentration, and  $C_n$  the concentration of the ion  $H_nP$ , we may write,

$$\begin{aligned} C_n &= (H^+) \gamma_{n-1} C_{n-1} / \gamma_n K_{2\nu-n+1} \\ C^0 &= \sum_{n=0}^{2\nu} C_n \end{aligned} \quad (38)$$

where  $\gamma_n$  is the activity coefficient of the ion  $H_nP$ . The difference equations 38 for the concentrations  $C_n$  have the solutions

$$\begin{aligned} C_n &= C^0 K_n (H^+)^n / G[(H^+)] \\ G[(H^+)] &= \sum_{k=0}^{2\nu} K_k (H^+)^k \end{aligned} \quad (39)$$

$$K_n = \gamma_n^{-1} \prod_{s=1}^n K_{2\nu-s+1}^{-1}$$

Thus the fraction  $f_n$  of the ampholyte existing in the form of the ion  $H_nP$  is

$$f_n = K_n (H^+)^n / G[(H^+)] \quad (40)$$

The mean charge  $\bar{z}$  of the ampholyte is evidently given by

$$\bar{z} = \sum_{n=0}^{2\nu} (n - \nu) f_n = \frac{d \log G}{d \log (H^+)} - \nu \quad (41)$$

The isoelectric point corresponding to vanishing average charge is obtained by solving the algebraic equation

$$\frac{d \log G}{d \log (H^+)} - \nu = 0 \quad (42)$$

for  $(H^+)$ , equal to  $10^{-pH}$ . The activity coefficient of the ampholyte, regarded as the neutral species  $H\nu P$ , is  $f_\nu \gamma_\nu$ , which may be written

$$\begin{aligned} \gamma &= \frac{\gamma_\nu}{1 + \gamma_\nu \phi[(H^+)]} \\ \phi[(H^+)] & \end{aligned} \quad (43)$$

In an aqueous solution containing the ampholyte and a strong acid at a concentration  $C_A$ , electrical neutrality requires

$$\sum_{n=0}^{2\nu} (n - \nu) f_n C^0 + C_{H^+} - C_A - C_{OH^-} = 0. \quad (44)$$

The bound hydrogen ion per mole of ampholyte,  $h$ , is defined as follows

$$C_{H^+} = C_A + C_{OH^-} - C^0 h \quad (45)$$

and the bound hydrogen ion becomes equal to the average charge,  $\bar{z}$ , or

$$h = \frac{d \log G}{d \log (H^+)} - \nu \quad (46)$$

All pertinent information concerning the acid-base equilibria of the ampholyte may thus be calculated from the function  $G[(H^+)]$

$$G[(H^+)] = \sum_{n=0}^{2\nu} K_n (H^+)^n \quad (47)$$

$$K_n = \gamma_n^{-1} \prod_{s=1}^n K_{2\nu-s+1}^{-1}$$

We shall now undertake the development of an approximate statistical theory of the function  $G[(H^+)]$  for a large spherical ampholyte of radius  $R$ , containing  $\nu$  acidic groups  $\text{COOH}$  and  $\nu$  basic groups  $\text{NH}_3^+$ . Let us suppose that  $K_1^0$  and  $K_2^0$  are the respective dissociation constants of an isolated  $\text{COOH}$  group and an isolated  $\text{NH}_3^+$  group in the absence of electrostatic interaction between the protons of the ampholyte ion. In the ion  $\text{H}_n\text{P}$ , the  $n$  protons can assume a number of configurations corresponding to the different ways in which the  $\nu$  basic sites  $\text{COO}^-$  and the  $\nu$  basic sites  $\text{NH}_2$  may be assigned to them. If  $W_c$  is the local free energy of the configuration  $C$ , the configurational part of the chemical potential  $\mu^0 n$  of the ion HP may be computed from the partition functions.

$$e^{-(\mu_n^0 - \mu_0^0 - \nu \mu_{H^+}^0)/kT} = \sum_C e^{-W_c/kT} \quad (48)$$

$$W_c = n_1 kT \log K_1^0 + n_2 kT \log K_2^0 + V_c$$

where  $n_1$  is the number of protons occupying  $\text{COO}^-$  sites,  $n_2$  the number occupying  $\text{NH}_2$  sites, and  $V_c$  is the mutual electrostatic energy of the  $n$  protons and of each proton with the negative charges of the  $\text{COO}^-$  groups other than that occupied by it. The calculation of  $V_c$  demands a specification of the location of the basic groups on the surface of the molecule and use of the Kirkwood-Westheimer theory. We shall assume that the basic sites are randomly distributed on the surface of the sphere  $R$ . On the

pair of charges averaged over all points on the surface of the sphere is simply  $e_1 e_2 / 2DR$ , where  $D$  is the dielectric constant of the solvent. The average electrostatic energy may therefore be roughly approximated by

$$\bar{V} = [n(n-1) - 2n_1(\nu-1) - 2n_2\nu] \epsilon^2 / 2DR \quad (49)$$

If fluctuations in  $V_e$  are neglected in the calculation of the partition function 48 we may now write

$$\begin{aligned} e^{-[\mu_n^0 - \mu_0^0 - \nu\mu_{H^+}^0] / kT} &= e^{\nu^2\alpha - (n-\nu)^2\alpha} A_n \\ \alpha &= \epsilon^2 / 2DRkT \\ A_n &= \sum_{\substack{n_1, n_2 \\ n_1+n_2=n}}^n \binom{\nu}{n_1} \binom{\nu}{n_2} \lambda_1^{n_1} \lambda_2^{n_2} \quad (50) \\ \lambda_1 &= (K_1^0 e^\alpha)^{-1}; \quad \lambda_2 = (K_2^0 e^{-\alpha})^{-1} \end{aligned}$$

since  $\binom{\nu}{n_1} \binom{\nu}{n_2}$  is the number of configurations corresponding to  $n_1$  protons on  $\text{COO}^-$  sites and  $n_2$  on  $\text{NH}_2$  sites. By thermodynamics, we have

$$kT \log K_{2\nu-n+1} = \mu_n^0 - \mu_{n-1}^0 - \mu_{H^+}^0 \quad (51)$$

and from equations 50 and 51 we may write

$$K_{2\nu-n+1} = \frac{A_{n-1}}{A_n} e^{[2(n-\nu)-1]\alpha} \quad (52)$$

Thus the coefficient  $K_n$  in the function  $G[(H^+)]$  becomes

$$K_n = \gamma_n^{-1} \prod_{s=1}^{\nu} K_{2\nu-s+1}^{-1} = e^{\nu^2\alpha - (n-\nu)^2\alpha - \log \gamma_n} A_n \quad (53)$$

In all calculations the factor  $e^{\nu^2\alpha}$  is of no importance, so that it is sufficient to write

$$\begin{aligned} G[(H^+)] &= \sum_{n=0}^{2\nu} e^{-(n-\nu)^2\alpha - \log \gamma_n} A_n (\lambda_1, \lambda_2) [(H^+)]^n \\ A_n (\lambda_1, \lambda_2) &= \text{Coefficient of } t^n \text{ in } (1 + \lambda_1 t)^\nu (1 + \lambda_2 t)^\nu \end{aligned} \quad (54)$$

We shall not discuss in detail the electrostatic effect embodied in the factor  $e^{-(n-\nu)^2\alpha}$ , except to remark that it favors the dipolar ion form of the ampholyte H<sub>2</sub>P. However, due to the fact that  $\alpha$  may be small in a protein molecule, the electrostatic interaction between the protons of the ampholyte may be dominated by the interaction with the electrolytic environment, manifesting itself in the  $\log \gamma_n$  term. Thus acid-base equilibrium and the isoelectric point may be expected to be sensitive to the ionic strength of

approximation to  $G(\text{H}^+)$  with the neglect of electrostatic interaction within the ampholyte molecule and with the environment.

$$G(\text{H}^+) = (1 + (\text{H}^+)/K_1^0)^{\nu} (1 + (\text{H}^+)/K_2^0)^{\nu} \quad (55)$$

By equation 41, the mean charge of the ampholyte becomes

$$\bar{z} = \nu \left\{ \frac{(\text{H}^+)}{(\text{H}^+) + K_1^0} + \frac{(\text{H}^+)}{(\text{H}^+) + K_2^0} - 1 \right\} \quad (56)$$

For the isoelectric point,  $\bar{z} = 0$ , we have

$$(\text{H}^+) = (K_1^0 K_2^0)^{1/2} \quad (57)$$

a value independent of the ampholyte's specific structure. This corresponds to a pH of approximately 6. The specific behavior of an ampholyte is thus to be attributed to electrostatic interaction between its charges and with its electrolyte environment. The theory may be extended without difficulty to an ampholyte containing acidic groups other than  $\text{NH}_3^+$  and  $\text{COOH}$ , for example,  $\text{NH}_2^+$  and  $\text{SH}$  groups.

### Dielectric Polarization

The large magnitude of the electric moments of dipolar ions should, according to the theory of Debye<sup>22</sup> manifest itself in the dielectric constants of their solutions. This prediction has been confirmed by Wyman, Devoto and others<sup>8</sup>, who found that the aliphatic amino acids and peptides produce remarkably large dielectric constant increments in aqueous solution and mixed polar solvents. However, attempts to interpret the measurements of these investigators on the basis of the Lorentz field and the Clausius-Mosotti formula failed completely. Wyman devised an empirical scheme for computing the electric moments of dipolar ions from the dielectric constant increments, which he deduced from a study of the dielectric polarization of mixtures of simple polar molecules of known moment. Wyman's scheme is qualitatively confirmed by Onsager's theory (11) of the dielectric polarization of polar liquids, which replaces the Lorentz local field with a much more satisfactory approximation. As an extension of Onsager's ideas, Kirkwood (12) has developed a general statistical theory of the dielectric polarization of polar liquids and polar mixtures. In the latter theory, it becomes apparent that the dipole moment  $\mu$  of an individual molecule cannot be obtained from dielectric constant measurements alone. However, a moment  $\bar{\mu}$ , equal to  $\sqrt{\mu\bar{\mu}}$ , may be calculated in a straightforward manner, the moment  $\bar{\mu}$  being the sum of the electric moment of a molecule and the moment which it induces in its neighboring molecules by hindering their rotation relative to itself.

As we shall show, this appears to be true of the dipolar ions of the aliphatic amino acids in dilute aqueous solution.

The Onsager-Kirkwood equation for a binary polar mixture has the form,

$$\begin{aligned}\frac{D-1}{3}v &= \frac{3D}{2D+1} \{N_1 P_1 + N_2 P_2\} \\ P_1 &= \frac{4\pi N}{3} [\alpha_1 + \mu_1 \bar{\mu}_1 / 3kT] \\ P_2 &= \frac{4\pi N}{3} [\alpha_2 + \mu_2 \bar{\mu}_2 / 3kT]\end{aligned}\quad (58)$$

where  $D$  is the dielectric constant,  $v$  the mean molar volume,  $N_1$  and  $N_2$  the mole fractions of the components and  $\alpha_1$  and  $\alpha_2$  their optical polarizabilities. The moments  $\mu_1$  and  $\mu_2$  are the molecular dipole moments, while  $\bar{\mu}_1$  and  $\bar{\mu}_2$  are sums of the corresponding molecular dipole moments and those induced by hindered rotation in the molecular environments. For simplicity, we assume  $\mu_2$  and  $\bar{\mu}_2$  to be parallel vectors, in the present discussion. It is convenient to define a molar dielectric constant increment,  $\delta$ , by the relation

$$D = D_1 + \delta C_2 \quad (59)$$

where  $C_2$  is the molar concentration of the solute component and  $D_1$  is the dielectric constant of the pure solvent. If  $D$  is large, of the order of magnitude of 100, the factor  $3D/(2D+1)$  in equation 58 may be adequately approximated by  $3/2$ . With this approximation, equations 58 and 59 yield

$$\delta = \frac{9}{2000} \left\{ P_2 - \frac{\varphi_2 P_1}{v_1} + \frac{v}{v_1} \frac{P_1 - P_1^0}{N_2} \right\} \quad (60)$$

where  $v_1$  is the molar volume of the pure solvent,  $P_1^0$ , its polarization, and  $\varphi_2$  is the apparent molar volume of the solute. At infinite dilution, we may write

$$\delta_0 = \frac{9}{2000} \left\{ P_2^0 - \frac{\varphi_2^0}{v_1} P_1^0 + \left( \frac{\partial P_1}{\partial N_2} \right)_0 \right\} \quad (61)$$

Wyman has found that  $\delta$  is approximately constant in solutions of the aliphatic amino acids and peptides up to concentrations of the order of one mole per liter. This is in accord with the theory, since  $P_2$ , the dominant term in  $\delta$ , will not be expected to deviate much from  $P_2^0$  as long as the mole fraction of the solute remains small. It is true that the difference  $\bar{\mu} - \mu$  must depend on dipole-dipole interaction of the solute molecules with each other as well as with the solvent molecules and that  $P_2$  may be



If we define  $\bar{\mu}_2$  as the geometric mean  $\sqrt{\mu_2 \bar{\mu}_2}$ , and introduce numerical values for the universal constant, equations 58 and 61 yield at 25°

$$\bar{\mu}_2 = 3.30 \left\{ \delta_0 + .0045 \left( \frac{\varphi_2^0 P_1^0}{v_1} - \left( \frac{\partial P_1}{\partial N_2} \right)_0 - R_2 \right) \right\}^{1/2} \quad (62)$$

where  $R_2$  is the molar refraction of the solute and  $\bar{\mu}_2$  is expressed in Debye units. For dipolar ions the last three terms in the parentheses are not large, so that to a fair approximation, we may write,

$$\bar{\mu}_2 = 3.30 \delta_0^{1/2} \quad (63)$$

Wyman's value of  $\delta_0$  for glycine in water, 22.6, leads to a moment  $\mu_2$  of 15.7 Debye units. The dipole moment  $\mu_2$  has the value 15.0 from salting-in data. From  $\bar{\mu}_2$  and  $\mu_2$ , we then estimate  $\bar{\mu}_2$  as 16.4 Debye units. It appears that  $\bar{\mu}_2$  differs from  $\mu_2$  by only ten per cent. The difference 1.4 represents the vector sum of the moments induced in the solvent environment through orientation of adjacent water molecules by the positive and negative groups of the dipolar ion. Since this additional moment turns out to be a small fraction of the total moment of the dipolar ion, we may conclude that  $\bar{\mu}_2$  is a rather good approximation to  $\mu_2$ , the dipolar ion moment. The effective dipole separation  $\bar{\mu}_2/\epsilon$  corresponding to  $\bar{\mu}_2$  is 3.27 Å, in good agreement with the structural value, 3.17 Å. In Table 9, values of  $\bar{\mu}_2$  and the corresponding dipole separations  $R$ , computed from the  $\delta$  values of Wyman and Devoto, are presented<sup>23</sup>. The dipole distances,  $R$ , computed from the dielectric constant increments lie very close to the mean distances between the  $\text{NH}_3^+$  and  $\text{COO}^-$  groups of the respective dipolar ions, calculated on the assumption of free rotation around carbon-carbon bonds. They are considerably smaller than in an extended chain configuration. We may conclude that the dielectric polarization of solutions of the amino acids and peptides amply confirms the dipolar ion hypothesis and leads to structural parameters in essential agreement with those calculated from other properties.

Information concerning the rotatory diffusion constants and the size and shape of polar molecules may be obtained from dielectric loss measurements in the region of anomalous dispersion. Such measurements can be interpreted by means of the Debye<sup>22</sup> theory of anomalous dispersion and the extended Onsager theory of the local field. The investigation of the smaller amino acids and peptides in solvents of low viscosity such as water or alcohol is inconvenient, since the region of anomalous dispersion generally lies well on the short wave length side of the radio frequency range. In solutions of proteins, on the other hand, the anomalous dispersion lies within a frequency range accessible to measurement with radio-

tions of the proteins have been carried out by Oncley<sup>9</sup>, Ferry<sup>9</sup>, Williams<sup>10</sup> and others. (See Chapter 22.)

Although a complete theory of anomalous dispersion in polar liquids has not yet been developed, certain approximations adequate for solutions of components of widely different relaxation times may be stated. On the Debye theory, an assembly of dipole molecules diffuse to their equilibrium configuration in an applied electric field by rotatory Brownian motion. In an alternating field of sufficiently high frequency, the rate of Brownian motion becomes insufficient to permit realization of equilibrium with the field and a phase difference is set up between the field and the polarization due to dipole orientation. The associated displacement current  $4\pi \frac{\partial P}{\partial t}$  acquires a conductance as well as a reactance component, and dielectric loss sets in. This situation may be described by means of a complex dielectric constant  $D' - iD''$  where  $\omega D'E$  is the reactance component of the current and  $\omega D''E$  is the conductance component in a field  $E$  of frequency  $\omega/2\pi$ . Thus, when a solution of polar molecules is subjected to an alternating electric field,  $E_0 e^{i\omega t}$ , of frequency  $\omega/2\pi$ , the polarizations,  $P_1$  and  $P_2$  of equation 58, become complex quantities:

$$P_1 = P'_1 - iP''_1$$

$$P_2 = P'_2 - iP''_2$$

leading to a complex dielectric constant  $D' - iD''$  and a loss angle  $\tan^{-1}(D''/D')$ . We shall assume that intermolecular relaxation effects can be neglected, that is to say, that  $\bar{\mu}_1$  and  $\bar{\mu}_2$  of equation 58 are independent of frequency. Under this assumption, we may write for a rigid ellipsoidal molecule

$$P = \frac{4\pi N}{3} \left\{ \alpha_1 + \frac{1}{3kT} \left[ \frac{\mu_a \bar{\mu}_a}{1 + i\omega\tau_a} + \frac{\mu_b \bar{\mu}_b}{1 + i\omega\tau_b} + \frac{\mu_c \bar{\mu}_c}{1 + i\omega\tau_c} \right] \right\} \quad (64)$$

where  $\mu_a, \mu_b, \mu_c$ , and  $\bar{\mu}_a, \bar{\mu}_b, \bar{\mu}_c$ , are the components of the electric moment  $\mu$  and  $\bar{\mu}$  in the directions of the axes  $a, b, c$  of the ellipsoid, and  $\tau_a, \tau_b, \tau_c$  are relaxation times associated with molecular rotations around axes perpendicular to  $a, b, c$ . If  $\Theta_a, \Theta_b, \Theta_c$  are the corresponding rotatory diffusion constants, the generalized Debye theory yields

$$\tau_a = \frac{1}{\Theta_b + \Theta_c}$$

$$\tau_b = \frac{1}{\Theta_a + \Theta_c} \quad (65)$$

By the theory of rotatory Brownian motion, each rotatory diffusion constant, for example  $\Theta_a$ , is related to the corresponding hydrodynamic resistance  $\zeta_a$  as  $kT/\zeta_a$ . Perrin<sup>24</sup> has investigated the hydrodynamic problem relating to the rotation of an ellipsoid in a viscous medium. His results are used to calculate the lengths of the molecular axes from the measured rotatory diffusion constants and relaxation times.

The equations corresponding to equation 59 in a frequency range in which the anomalous dispersion of the solvent can be neglected are

$$\begin{aligned} D' &= D_1 + \delta' C_2 \\ D'' &= \delta'' C_2 \end{aligned} \quad (66)$$

$$\begin{aligned} \delta' &= \frac{9}{2000} \left\{ \frac{P_{2a}}{1 + \omega^2 \tau_a^2} + \frac{P_{2b}}{1 + \omega^2 \tau_b^2} + \frac{P_{2c}}{1 + \omega^2 \tau_c^2} \right\} \\ \delta'' &= \frac{9\omega}{2000} \left\{ \frac{\tau_a P_{2a}}{1 + \omega^2 \tau_a^2} + \frac{\tau_b P_{2b}}{1 + \omega^2 \tau_b^2} + \frac{\tau_c P_{2c}}{1 + \omega^2 \tau_c^2} \right\} \end{aligned}$$

where  $D'$  and  $D''$  are the real and imaginary parts of the dielectric constant and terms depending upon the volume displacement of the solvent and the optical polarization of the solute have been neglected. Since the application of these equations is described in Chapter 22, we shall not present further details here. In Oncley's analysis of dielectric polarization data for proteins, he employed equations for  $\delta'$  and  $\delta''$  differing from equation 66 only by a numerical factor fixed to yield the structural dipole separation 3.17 Å for glycine, and he interprets  $\bar{\mu}$  as the solute dipole moment rather than as  $\sqrt{\mu\bar{\mu}}$ .

The influence of molecular interaction does not appear to be important in solutions of the proteins, but it presents a theoretical problem which demands further study. It should also be borne in mind that the interpretation of loss data on the basis of Perrin's rigid ellipsoidal model may sometimes be open to question. If the molecule possesses internal rotatory degrees of freedom upon which its dipole moment depends, a distribution in relaxation times associated with internal rotatory Brownian motion is to be expected. Internal dispersion of this type is possibly important in a linear protein such as zein, though probably unimportant for an approximately spherical protein such as egg albumin.

#### Appendix to Section 2

We shall present here some of the mathematical details in the calculation of  $V_{ik}$ , the electrostatic work required to bring a dipolar ion  $i$  and a real ion  $k$  from infinite separation in the pure solvent to a given relative con-

polar ion by a charge distribution  $e_1 \cdots e_s$  located in a cavity  $\omega_0$ , of dielectric constant  $D_i$ , in the solvent of dielectric constant  $D$ . The potential  $V_{ik}$  is then given by

$$V_{ik} = W - W_0 \quad (67)$$

where  $W$  is the work of charging the system in the given configuration and  $W_0$  the work of charging when the two ions are infinitely separated.  $W$  is to be calculated by means of the formula

$$W = 1/2 \left\{ z_k e \psi_e(\mathbf{r}_k) + \sum_{l=1}^s e_l \psi_i(\mathbf{r}_l) \right\} \quad (68)$$

where  $\psi_i(\mathbf{r}_l)$  is the electrostatic potential in the interior of  $\omega_0$  at the point  $\mathbf{r}_l$  of location of charge  $e_l$ , and  $\psi_e(\mathbf{r}_k)$  is the potential exterior to  $\omega_0$  at the point of location  $\mathbf{r}_k$  of the real ion. The potentials  $\psi_i$  and  $\psi_e$  both satisfy Laplace's equation:

$$\nabla^2 \psi = 0 \quad (69)$$

as well as the boundary conditions:

$$\psi_i(\mathbf{r}) = \psi_e(\mathbf{r}) \quad (70)$$

$$D_i \mathbf{n} \cdot \nabla \psi_i(\mathbf{r}) = D \mathbf{n} \cdot \nabla \psi_e(\mathbf{r})$$

on the surface of the cavity  $\omega_0$ .

Sphere: When the cavity  $\omega_0$  is a sphere of radius  $b$ , it is convenient to employ polar coördinates  $(r, \theta, \phi)$  with origin at the center of the sphere. Potentials  $\psi_i$  and  $\psi_e$  satisfying Laplace's equation and possessing the appropriate singularities are

$$\begin{aligned} \psi_i &= \sum_{l=1}^s \frac{e_l}{D_i |\mathbf{r} - \mathbf{r}_l|} + \sum_{n=0}^{\infty} \sum_{m=-n}^{+n} K_{Enm} r^n P_n^m(\cos \theta) e^{im\phi} \\ \psi_e &= \frac{z_k e}{D |\mathbf{r} - \mathbf{r}_k|} + \sum_{n=0}^{\infty} \sum_{m=-n}^{+n} A_{nm} r^{-n-1} P_n^m(\cos \theta) e^{im\phi} \end{aligned} \quad (71)$$

where the  $P_n^m(\cos \theta)$  are the associated Legendre functions of the first kind. On the boundary of the sphere  $b$ , we have:

$$\psi_e(b, \theta, \phi) = \psi_i(b, \theta, \phi)$$

INSTITUTE OF SC  
(72)

for all values of  $\theta$  and  $\phi$  in the intervals 0 to  $\pi$  and 0 and  $2\pi$ . On the surface of the sphere, we may employ the following harmonic expansions:

$$\begin{aligned} \frac{z_k \epsilon}{D |\mathbf{r} - \mathbf{r}_k|} &= \sum_{n=0}^{\infty} \sum_{m=-n}^{+n} F_{nm} r^n P_n^m(\cos \theta) e^{im\phi} \\ F_{nm} &= \frac{z_k \epsilon}{Dr_k^{n+1}} \frac{(n - |m|)!}{(n + |m|)!} P_n^m(\cos \theta_k) e^{-im\phi} \\ \sum_{i=1}^s \frac{e_i}{D_i |\mathbf{r} - \mathbf{r}_i|} &= \sum_{n=0}^{\infty} \sum_{m=-n}^{+n} G_{nm} r^{-n-1} P_n^m(\cos \theta) e^{im\phi} \\ G_{nm} &= \frac{1}{D_i} \frac{(n - |m|)!}{(n + |m|)!} \sum_{i=1}^s e_i r_i^n P_n^m(\cos \theta) e^{im\phi}. \end{aligned} \quad (73)$$

Substitution of equations 71 and 73 in equation 72 and use of the orthogonality of the functions,  $P_n^m(\cos \theta) e^{im\phi}$  on the surface of the sphere, we obtain the following set of linear equations for the coefficients  $A_{nm}$  and  $K_{Rnm}$ :

$$\begin{aligned} A_{nm} + b^{2n+1} F_{nm} &= G_{nm} + b^{2n+1} K_{Rnm} \\ (n+1)A_{nm} - nb^{2n+1} F_{nm} &= \sigma[(n+1)G_{nm} - nb^{2n+1} K_{Rnm}] \\ \sigma &= D_i/D. \end{aligned} \quad (74)$$

Solution yields

$$\begin{aligned} K_{Rnm} &= \frac{2n+1}{n+1+n\sigma} F_{nm} + \frac{(n+1)(\sigma-1)}{n+1+n\sigma} \frac{G_{nm}}{b^{2n+1}} \\ A_{nm} &= \frac{n(1-\sigma)b^{2n+1}}{n+1+n\sigma} F_{nm} + \frac{(2n+1)\sigma}{n+1+n\sigma} G_{nm}. \end{aligned} \quad (75)$$

Use of equations 71 and 75 in equations 67 and 68 and application of the addition theorem of spherical harmonics yields when the dipolar ion has no net charge

$$\begin{aligned} V_{ik} &= \frac{z_k \epsilon}{Dr_k} \sum_{n=1}^{\infty} \sum_{l=1}^s \frac{2n+1}{n+1+n\sigma} \frac{e_l r_l^n}{r_k^n} P_n(\cos \theta_{kl}) \\ &+ \frac{z_k^2 \epsilon^2}{2Dr_k^2} b(1-\sigma) \sum_{n=1}^{\infty} \frac{n}{n+1+n\sigma} \left(\frac{b}{r_k}\right)^{2n} \end{aligned} \quad (76)$$

where  $r_k$  is the distance of the real ion  $z_k \epsilon$  from the center of the sphere and  $\theta_{kl}$  is the angle between the vectors  $\mathbf{r}_k$  and  $\mathbf{r}_l$  from the center of the sphere terminating in the real ionic charge  $z_k \epsilon$  and the charge  $e_l$  of the dipolar ion. When the dipolar ion contains a point dipole at the center, the first sum in (76) degenerates into

and equation 11 results at once from equation 76, with a slight change in the summation index in the second sum. In this case the dipolar ion may be regarded as possessing two charges,  $+\epsilon$  and  $-\epsilon$ , situated at equal distances  $r_0$  from the center, with  $\theta_{k2} = \pi - \theta_{k1}$ . If we pass to the limit  $r_0 = 0$  with  $\mu = 2 \lim_{r_0 \rightarrow 0} (\epsilon r_0)$  all terms except for  $n = 1$  vanish in first sum of equation 76, the first term reducing to 77, if  $\theta_{ki}$  is designated simply by  $\theta_k$ .

Before leaving the spherical case, it is perhaps desirable to give the general expression for  $K_{Ri}^{(0)}$  when the charge distribution of the dipolar ion is arbitrary.  $K_{Ri}^{(1)}$  is still given by equation 42.

$$K_{Ri}^{(0)} = \frac{4\pi N \epsilon^2}{2303 (DkT)^2} \sum_{n=1}^{\infty} \frac{2n+1}{(2n-1)(n+1+n\sigma)} \frac{M_n}{\alpha^{2n-1}} \quad (78)$$

$$M_n = \sum_{l,l'=1}^s e_l e_{l'} r_l^n r_{l'}^n P_n(\cos \theta_{ll'})$$

where  $\theta_{ll'}$  is the angle between the vectors of length  $r_l$  and  $r_{l'}$ , joining the charges  $e_l$  and  $e_{l'}$  to the center of the sphere. For a molecule of the type of cystine, a model consisting of two point dipoles of moment  $\mu$  perpendicular to a common diameter and each situated at a distance  $l$  from the center, is useful. In this case equation 78 reduces to

$$K_{Ri}^{(0)} = \frac{8\pi N \epsilon^2}{2303 (DkT)^2} \frac{\mu^2 (1 + \cos \theta)}{a} \sum_{n=0}^{\infty} \frac{2n+3}{(2n+1)[n+2+(n+1)\sigma]} \left(\frac{l}{a}\right)^{2n} \quad (79)$$

where  $\phi$  is the angle between the two dipole moments.

*Ellipsoid:* When the cavity  $\omega_0$  is ellipsoidal in form, we may conveniently employ confocal elliptical coordinates,  $\lambda, \eta, \phi$ , where  $\lambda = (r_1 + r_2)/R$  and  $\eta = (r_1 - r_2)/R$ ,  $r_1$  and  $r_2$  being the distances of the point from the respective foci and  $R$  the interfocal distance. The angle  $\phi$  measures the inclination of the plane  $r_1$  and  $r_2$  to a chosen reference plane containing the major axis. The cavity  $\omega_0$  is then specified by a value  $\lambda_0$  equal to the reciprocal of the eccentricity of its elliptical section. We suppose the charges  $e_1 \cdots e_s$  of the dipolar ion to lie on the major axis of the ellipsoid. We shall further neglect the image distribution induced in the cavity  $\omega_0$  by the real ionic charge. Potentials satisfying Laplace's equation and having the proper singularities are the following

$$\psi_e = \sum_{n=0}^{\infty} A_n P_n(\eta) Q_n(\lambda)$$

$$\psi_i = \sum_{l=1}^s \frac{e_l}{D_i |\mathbf{r} - \mathbf{r}_l|} + \sum_{n=0}^{\infty} K_{Rn} P_n(\eta) P_n(\lambda) \quad (80)$$

The boundary conditions are

$$\left. \begin{aligned} \psi_e(\lambda_0, \eta) &= \psi_i(\lambda_0, \eta) \\ D \left[ \frac{\partial \psi_e}{\partial \lambda} \right]_{\lambda=\lambda_0} &= D_i \left[ \frac{\partial \psi_i}{\partial \lambda} \right]_{\lambda=\lambda_0} \end{aligned} \right\} -1 \leq \eta \leq +1. \quad (81)$$

On the boundary of  $\omega_0$ , the initial terms of the second of equations 80 may be developed in the Neumann expansion;

$$\begin{aligned} \sum_{i=1}^s \frac{e_i}{D_i |\mathbf{r} - \mathbf{r}_i|} &= \frac{2}{R} \sum_{n=0}^{\infty} (2n+1) G_n P_n(\eta) Q_n(\lambda_0) \\ G_n &= \sum_{i=1}^s e_i P_n(\eta_i) \end{aligned} \quad (82)$$

where  $(1, \eta_i)$  are elliptical coordinates of the dipolar ionic charge  $e_i$ . Application of the boundary conditions 80 to equations 80 and 82, and use of the orthogonality of the functions,  $P_n(\eta)$ , yield the set of linear equations,

$$\begin{aligned} A_n Q_n(\lambda_0) &= K_{Rn} P_n(\lambda_0) + \frac{2}{R} (2n+1) G_n Q_n(\lambda_0) \\ A_n Q'_n(\lambda_0) &= \sigma \left[ K_{Rn} P'_n(\lambda_0) + \frac{2}{R} (2n+1) G_n Q'_n(\lambda_0) \right] \end{aligned} \quad (83)$$

where  $Q'_n(\lambda_0)$  and  $P'_n(\lambda_0)$  are the first derivatives of the indicated functions, and  $\sigma$  is the ratio,  $D_i/D$ . Solutions of the equations 83 and elimination of the derivative  $Q'_n(\lambda_0)$  by means of the formula

$$P_n Q'_n - P'_n Q_n = (1 - \lambda^2)^{-1}$$

yield

$$\begin{aligned} A_n &= \frac{2\sigma}{R} \frac{(2n+1)G_n}{1 + (\lambda_0^2 - 1)(1 - \sigma)P'_n(\lambda_0)Q_n(\lambda_0)} \\ K_{Rn} &= \frac{2(\sigma - 1)}{R} \frac{Q_n(\lambda_0)}{P_n(\lambda_0)} \frac{(2n+1)G_n}{1 - \sigma [P'_n(\lambda_0)Q_n(\lambda_0)]/[P_n(\lambda_0)Q'_n(\lambda_0)]}. \end{aligned} \quad (84)$$

We now calculate  $V_{ik}$  by means of the formula

$$V_{ik} = z_k e \psi_e(\lambda_k, \eta_k) \quad (85)$$

and obtain, neglecting  $\sigma$  in comparison with unity in the denominators on the right-hand side of equation 84,

$$V_{ik} = \frac{2z_k e}{R} \sum_{n=0}^{\infty} (2n+1) G_n P_n(\eta_k) Q_n(\lambda_k). \quad (86)$$

The sum begins with  $n$  equal to unity, since  $G_0$ , the total charge of the dipolar ion, vanishes. By a simple algebraic transformation equation 86 may be written as follows

$$V_{ik} = \frac{2z_k \epsilon}{DR} \frac{1}{1 + (\lambda_0^2 - 1)Q_1(\lambda_0)} \sum_{n=1}^{\infty} (2n+1)G_n P_n(\eta_k) Q_n(\lambda_k) + Y(\lambda_k, \eta_k) \quad (87)$$

$$Y(\lambda_k, \eta_k) = \frac{2(\lambda_0^2 - 1)z_k \epsilon}{DR} \sum_{n=2}^{\infty} \frac{(2n+1)G_n [Q_1(\lambda_0) - P'_n(\lambda_0)Q_n(\lambda_0)] [P_n(\eta_k)Q_n(\lambda_k)]}{[1 + (\lambda_0^2 - 1)P'_n(\lambda_0)Q_n(\lambda_0)] [1 + (\lambda_0^2 - 1)Q_1(\lambda_0)]}$$

The term  $Y(\lambda_k, \eta_k)$  vanishes for the limiting eccentricities zero and unity of the ellipsoidal cavity  $\omega_0$ , and can be neglected without great error for intermediate eccentricities. We therefore have, approximately,

$$V_{ik} = \frac{2z_k \epsilon}{DR[1 + (\lambda_0^2 - 1)Q_1(\lambda_0)]} \sum_{n=1}^{\infty} (2n+1)G_n P_n(\eta_k) Q_n(\lambda_k) \quad (88)$$

When the charge distribution of the dipolar ion consists of two charges  $+e$  and  $-e$  situated at the foci,  $(1, 1)$  and  $(1, -1)$  respectively, we have

$$G_n = -e[(-1)^n - 1] \quad (89)$$

and the series in equation 88 may be summed to give equation 44, when we note that  $1 + (\lambda_0^2 - 1)Q_1(\lambda_0)$  is equal to  $\lambda_0 \left[ 1 - \frac{\lambda_0^2 - 1}{2} \log \left( \frac{\lambda_0 + 1}{\lambda_0 - 1} \right) \right]$ .

On the other hand, when the charge distribution consists of a point dipole of moment  $\mu$  at the focus  $(1, -1)$ , the  $G_n$  have the form

$$G_n = \lim_{x \rightarrow 0} \left[ P_n \left( \frac{2x}{R} - 1 \right) - (-1)^n \right]; \quad \mu = \lim_{x \rightarrow 0} (ex) \quad (90)$$

and calculation yields

$$G_n = (-1)^n n(n+1)\mu/R \quad (91)$$

Summation in equation 88 with the  $G_n$  of equation 91 yields equation 48.





## Part II

### The Size, Shape and Electric Charge of Protein Molecules



## Chapter 13

# The Structural Basis of the Protein Molecule: Evidence from Analysis and from the Action of Proteolytic Enzymes

BY JOHN T. EDSALL

In Part I of this book we have dealt with simple dipolar ions—amino acids, peptides, betaines and others—whose structures are exactly known. A more complex task confronts us in this and the subsequent chapters. Proteins are built principally, and apparently in some instances entirely, of amino acids bound together in peptide linkage. Much of the knowledge set forth in Part I may therefore be employed to draw fairly direct inferences as to the nature of proteins. It may be inferred, for instance, that proteins at their isoelectric points are generally dipolar ions, carrying negatively charged carboxyl groups, and positively charged ammonium groups (of lysine), guanidinium groups (of arginine), and sometimes imidazolium groups (of histidine). It may also be expected that their solutions should have high dielectric constants, depending on the arrangement of these charged groups in the molecule. These inferences are abundantly verified by experiment.

### Some General Characteristics of Protein Molecules

The problem of the proteins, however, is vastly more complex than that of the simple molecules considered in Part I. Of all types of molecules found in nature, proteins are probably the most varied, the most complex, and the largest. For no protein has any detailed and convincing proof of structure been given. The great size of the proteins, and the number of different amino acids they yield on hydrolysis, give scope for an almost endless variety of possible structures. Furthermore, many proteins contain other components than amino acids as an integral part of the molecule. These prosthetic groups may be carbohydrate in nature, such as the polysaccharide isolated from crystalline egg albumin<sup>1</sup>, containing mannose, glucosamine and another nitrogenous compound; they may be heme deriva-

held in ester linkage to a hydroxyamino acid, as in casein<sup>4</sup>. The copper-containing prosthetic group in the hemocyanins<sup>5, 6</sup> is another example.

In the nucleoproteins we find a class of compounds of the utmost importance, in which amino acid residues and nucleic acids are both present in comparable proportions. These few examples, chosen out of a vast number, indicate that protein chemistry is far more than the chemistry of amino acids and their possible combinations with one another. Nevertheless, in our subsequent discussion, we shall lay stress on the amino acids and their combinations, for this aspect of the problem of protein chemistry is the one specific to proteins; and in itself it raises endless problems that are still unsolved.

For the most part, in our discussion, we shall make no attempt to picture the structure of proteins in any detail, for the experimental basis for any such picture is still totally inadequate. There is, however, detailed evidence as to the size of many proteins, and considerable although less satisfactory evidence as to their shapes. From analytical data, electromotive force and electrophoretic measurements, much has been learned as to the nature and number of the electrically charged groups in proteins; and dielectric constant measurements have begun to furnish some evidence as to the arrangement of these groups. In subsequent chapters, the methods available for the attack on all these problems, and the results already obtained, will be considered. In the present chapter we shall consider the chemical evidence as to the mode of linkage of amino acids in the protein molecule. Fortunately, in spite of the complexity of protein structure, there is much that points to an underlying simplicity in the type of pattern common to many proteins of diverse types.

In the first place, all amino acids derived from proteins are  $\alpha$ -amino acids, except proline and hydroxyproline, which are cyclic  $\alpha$ -imino acids. Furthermore, it is nearly certain that all of the naturally occurring optically active amino acids in proteins have the same steric configuration<sup>7</sup>; that is,

the groups  $\begin{array}{c} \text{H} \quad \text{R} \\ \diagdown \quad \diagup \\ \text{C} \\ \diagup \quad \diagdown \\ \text{H}_3\text{N}^+ \quad \text{COO}^- \end{array}$  have invariably the same spatial relation

to each other. The synthetic work of Emil Fischer<sup>8</sup> brought forth much powerful evidence for the importance of the peptide bond as the principal

<sup>4</sup> Rimington, C., *Biochem. J.*, **21**, 1179, 1187 (1927).

<sup>5</sup> Peacock, A. C., *Biochem. Rev.*, **9**, 175 (1934).

<sup>6</sup> Klotz, H. M., *Biochem. J.*, **299**, 32 (1938).

<sup>7</sup> For a review of some of the evidence on this point, see Clarke, H. T., in "Organic Chemistry," edited by H. G. H. N. York, 1938, Vol. II, Chapter 10, which gives a most valuable survey of amino acid chemistry. See also an important series of papers by Klotz, B., and his co-workers.

link involved in joining one amino acid to another in the protein molecule. Not only were a number of simple peptides isolated from the enzymatic hydrolysis products of many proteins, but the synthetic polypeptides made by Fischer were readily hydrolyzed by enzyme extracts from the digestive tract. Later investigations have confirmed and greatly extended the evidence for the importance of the peptide linkage. It is in accord with the peptide theory<sup>9, 10</sup> that the enzymatic hydrolysis of proteins is very commonly found to involve the liberation of equivalent quantities of amino and carboxyl groups. The only exceptions hitherto found to this rule have been in proteins containing proline or oxyproline held in peptide ( $-\text{CO}\cdot\text{N}=\text{}$ ) linkage, in which case imino rather than amino groups are liberated on hydrolysis<sup>11</sup>.

It should be pointed out that the titrations on which these conclusions are based are not extremely accurate, the probable error being always at least 1%, and generally 3% or more. More decisive is the recent evidence, obtained by Bergmann and his collaborators, that peptide linkages in certain synthetic substrates are split by the typical proteolytic enzymes. Suitable substrates have now been found for no less than six such enzymes—crystalline trypsin and chymotrypsin, crystalline pepsin, liver cathepsin, papain and bromelin. The nature of the substrates and the peptide linkages which are split are listed in detail in Table 1.

#### Splitting of Synthetic Substrates by Proteolytic Enzymes; the Specificity of Proteinases

Unlike the simpler peptidases, which attack only terminal peptide linkages, these enzymes "split their substrates preferably, although not exclusively, at interior peptide linkages; they are endopeptidases. . . . The proteinases are particularly sensitive to the presence and the special nature of the side chains in the substrate; for example, papain distinguishes between glycyl and leucyl, and chymotrypsin between phenylalanyl and leucyl. . . . However, the splitting of a peptide linkage by a proteinase is not only affected by the two amino acid residues which directly participate in the peptide bond but by more distant groups as well; this effect is clearly visible in the decrease of splitting of the tyrosyl-glycine linkage by chymotrypsin on passing from carbobenzoxytyrosylglycineamide to carbobenzoxytyrosylglycylglycineamide" (reference 12, p. 407-408).

The difference in specificity between Papain I and chymotrypsin is well revealed by a study of the substrate carbobenzoxyglycyltyrosylglycineamide: carbobenzoxyglycyl | tyrosylglycine | amide (splitting by Papain I);

<sup>9</sup> Waldschmidt-Leitz, E. *Physiol. Rev.* 11, 250 (1931).

carbobenzoxylglycyltyrosyl | glycineamide (splitting by chymotrypsin). The vertical lines indicate the points of splitting by the two enzymes.

The splitting of such simple peptide linkages by proteolytic enzymes is decisive evidence (in the light of our knowledge of enzyme specificity) that peptide linkages of a fundamentally similar sort must exist within protein molecules.

TABLE 1. SOME PROTEOLYTIC ENZYMES AND THEIR SYNTHETIC SUBSTRATES  
Enzyme (reference in parentheses)                      Substrates Split

Crystalline trypsin (1) (6)	Bz·Ar   Am, Bz·G·Ly   Am; Bz·Ly   Am	
Crystalline chymotrypsin (2)	Cbz·T   G·Am Cbz·G·T   G·Am Cbz·G·Ph·G·Am	Cbz·T·G·G·Am G·T·G·Am
Crystalline pepsin (5)	Cbz·Gt   T Cbz·Gt·Ph Cbz·Ph·Gt	Cbz·Gt·T·Am Cbz·Gt   T·G
Papain I (3) (see also 1, 2)	Cbz·G   G Cbz·G   G·G Cbz·G   G·G·G Bz·G   Leu·G Bz·G   G·Leu·G Cbz·G   G·G   Leu·G Cbz·Leu·G   G Bz·Leu·G   G Bz·Leu   Leu·G Cbz·G   Gt·Am	Cbz·G   Gt·G Cbz·Gt   G Cbz·Gt   <sub>2</sub> G   <sub>1</sub> G G·G·Leu·G G·G·G·Leu·G Bz·G   Am Cbz·G   T·G   Am Bz·Ly   Am Bz·G   Ly·G Bz·G   <sub>2</sub> Cbz·Ly   <sub>1</sub> G
Cathepsin (Pig Liver) (4)	Cbz·G·Gt·G·Am Cbz·Leu·G·G	Cbz·G·G·G Cbz·G·G·G·Leu·G
Bromelin (Pineapple) (4)	Cbz·G·Gt·G·Am	Cbz·G·G·G·Leu·G

References: (1) Bergmann, M., Fruton, J. S., and Pollok, H., *Science*, 85, 410 (1937).

(2) Bergmann, M., and Fruton, J. S., *J. Biol. Chem.*, 118, 405 (1937).

(3) Bergmann, M., *Z. physikalische Chemie*, 115, 593 (1936).

(4) Bergmann, M., *J. Biol. Chem.*, 117, 627 (1939).

(5) Bergmann, M., *J. Biol. Chem.*, 130, 81 (1939).

In this table, Cbz = carbobenzoxy; Am = a terminal amide linkage; G = glycyl or glycine; Ar = Arginyl; Gt = Glutamyl; Bz = Benzoyl; Ly = Lysyl; Leu = Leucyl; Ty = Tyrosyl; Ph = Phenylalanyl.

The optically active amino acid residues are all the natural (l) forms. Thus Cbz·G·Ph·G·Am denotes carbobenzoxy-glycyl-l-phenyl-alanyl-glycine amide. Vertical lines are used to denote the point where the molecule is split by the given enzyme, when this point has been explicitly determined by isolation of the products. In two cases, the subscripts 1 and 2 are used to denote the preferential order of attack of a given enzyme on two different linkages in the same substrate.

These studies do not entirely exclude the possibility that diketopiperazine rings exist in proteins; the splitting of a peptide linkage in such a ring would, of course, give rise to equal numbers of amino and carboxyl groups. No satisfactory evidence has yet been forthcoming, however, that any synthetic diketopiperazine can be split by any enzyme found in nature<sup>13</sup>. The likelihood that such rings exist in proteins, therefore, appears small.

### The Nature of the Peptide Linkages and the Free Acid and Basic Groups of Proteins

Another important question arises. Are the peptide linkages in proteins always between  $\alpha$ -amino and carboxyl groups, or may they involve such groups as the distal carboxyl groups of aspartic or glutamic acid, the  $\epsilon$ -amino group of lysine, or the guanido group of arginine? The available evidence appears to be against the existence of any appreciable number of such links in proteins. We may summarize this evidence, largely following the treatment of Linderstrøm-Lang (reference 10, p. 434).

1. The  $\epsilon$ -amino group in lysine is probably free, for the treatment of proteins with nitrous acid liberates half the lysine nitrogen, or its equivalent<sup>14</sup>. Furthermore, it has been found that, in the hydrolysate from deaminized casein, lysine has disappeared from the basic fraction of the amino acids<sup>15</sup>. When proteins are benzoylated or benzoysulfonated,  $\epsilon$ -benzoyllysine and  $\epsilon$ -benzoysulfonyllysine are isolated after hydrolysis<sup>16</sup>.

2. The guanidino group of arginine is probably free.

On nitration of protamines, and subsequent hydrolysis of the nitrated protein, nitroarginine is obtained<sup>17</sup>. One protamine, clupein<sup>18</sup>, shows only one free amino group per 4000 gm of clupein—a value confirmed by the Van Slyke amino nitrogen determination—yet the total acid-binding capacity is more than twenty times as great as that to be expected from this amino group. The other acid-binding groups, therefore, must be the free guanidine groups of arginine residues. Moreover, it is found in general, on the enzymatic hydrolysis of proteins, that carboxyl and  $\alpha$ -amino groups are liberated in equivalent quantities—a fact which indicates that neither the guanidine group in arginine nor the imidazole group in histidine is involved in the linkages split.

3. One of the carboxyl groups of the dicarboxylic acids is either bound to ammonia or free. Glutamine and asparagine have been isolated from hydrolysates obtained by the action of enzymes upon edestin and gliadin<sup>19</sup>. Since the distal carboxyl groups are tied off in glutamine and asparagine, they could not have been free to combine with amino acids in peptide linkage. The state of dicarboxylic acid groups not bound as amide is not, of course, settled by this argument, but all available evidence is compatible with the view that the  $\alpha$ -carboxyl groups are bound in peptide linkage, while the distal carboxyl groups not bound as amide are free to ionize.

4. The hydroxyl groups in the oxyamino acids are probably free; except

<sup>14</sup> Van Slyke, D. D., and Birchard, F. J., *J. Biol. Chem.*, **16**, 539 (1913-14).

<sup>15</sup> Skraup, Z. H., and Kaas, K., *Ann.*, **351**, 379 (1907); Kossel, A., and Weiss, F., *Z. Physiol. Chem.*, **78**, 402 (1912).



in the phosphoproteins, in which some hydroxyl groups (especially those of serine, threonine and oxyglutamic acid) are bound in ester linkage to phosphoric acid groups<sup>4, 20</sup>. This has been inferred from the equivalence of carboxyl and  $\alpha$ -amino groups liberated during enzymatic digestion. The conclusion is probably correct, although, as we have already had occasion to point out, the titrations on which it is based are not extremely accurate.

Two additional lines of evidence indicate that the linkages in proteins are between carboxyl and  $\alpha$ -amino groups of amino acids. No enzyme has yet been found which will split any of the other conceivable types of linkage discussed above. Thus glutathione— $\gamma$ -glutamyl-cysteinyl-glycine—is split in one point, the cysteinyl-glycine linkage, by carboxypeptidase. No other peptidase or proteinase will attack either this split product or the original glutathione; the  $\gamma$ -glutamyl-cysteinyl linkage is resistant to all of them (reference 10, p. 469). Other evidence on this point, though relatively scanty, points in the same direction. Further study of a variety of synthetic substrates would be desirable to settle this point more definitely.

The other line of evidence comes from the x-ray diffraction studies discussed in Chapter 14. The data so obtained, at least for the fibrous proteins, are most naturally interpretable on the assumption of a regular succession of peptide linkages between  $\alpha$ -amino and carboxyl groups. Thus the cumulative weight of the evidence indicates that other types of linkage between amino acids, if present at all in proteins, are rare.

### The Protamines, and the Structure of Clupein

As yet, the detailed structure of no protein is known with anything approaching certainty. The class of proteins for which the structural pattern of the molecule is most nearly understood is probably that of the protamines<sup>21</sup>, those relatively small and strongly basic proteins which are obtained from the ripe sperm of certain fish. Within the cell, they appear to be combined with nucleic acids, and their extraction involves treatment with dilute sulfuric acid or cupric chloride. Necessarily, therefore, the protamines as isolated are decomposition products of the substances present in the spermatozoa, and the exact degree of decomposition in their preparation is unknown. The same may indeed be said of any extracted tissue protein. For the present, however, we may concentrate our attention on the properties of the protamines as isolated, since their study is an essential preliminary to the understanding of the larger complex present in the spermatozoön. Among the simplest and the best known of the

protamines are salmine (from salmon) and clupein (from herring). Both of these protamines contain approximately 88% of their nitrogen as arginine nitrogen. In the following discussion, though available data concerning salmine are similar, we shall deal chiefly with clupein, the most carefully studied of all the protamines. Clupein was studied analytically by Kossel and Dakin<sup>22</sup>, who isolated amino acids in the proportions: 2 valine, 1 serine, 1 alanine, 1 proline, 10 arginine. Practically identical results were reported by Waldschmidt-Leitz<sup>23</sup>, while Felix, Inouye and Dirr<sup>24</sup> reported the isolation of oxyproline. All workers agree that arginine is present in the ratio of two moles to one mole of monoamino (or imino) acid. There is clear evidence<sup>25</sup>, however, that clupein as ordinarily prepared is a mixture of different substances, very similar in character, but differing in the nature or arrangement of the monoamino acids present.

Mild acid hydrolysis (reference 21, p. 45) leads to the formation of tripeptides, known as protones, from simple protamines such as clupein and salmine. These contain two molecules of arginine (A) to one of monoamino (or imino) acid (M); their structure is probably: M—A—A, the free carboxyl group being on the terminal arginine. The evidence from enzymatic studies suggests that clupein is built up from a succession of such protones.

The application of proteolytic enzymes to elucidate the structure of protamines was begun by Kossel (reference 21, p. 40), and later carried on by Waldschmidt-Leitz and his coworkers, and by Felix and others<sup>26</sup>. Indeed, a fairly explicit formula for clupein was proposed by Waldschmidt-Leitz and Kofranyi<sup>27</sup> on the basis of degradation with a succession of enzymes specific for certain types of peptide linkage, and isolation of amino acids or peptides split off at each step. They concluded that their results were compatible only with the formula



where A denotes an arginyl residue, P a prolyl residue, and M a monoamino-monocarboxylic acid residue. (The order of the various monoamino acids remained undetermined.) This formula leads to a molecular weight of 2021 for clupein. The careful studies of Rasmussen and Linderstrøm-Lang<sup>18</sup>, however, led consistently to a minimal equivalent combining weight just twice as great as this. In any case, however, the molecule is extremely small as compared with most proteins.

<sup>22</sup> Kossel, A., and Dakin, H. D., *Z. Physiol. Chem.*, **41**, 407 (1904).

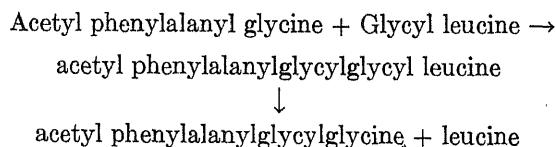
<sup>23</sup> Waldschmidt-Leitz, E., Ziegler, F., Schaffner, A., and Weil, L., *Z. Physiol. Chem.*, **197**, 219 (1931).

<sup>24</sup> Felix, K., Inouye, K., and Dirr, K., *Z. Physiol. Chem.*, **211**, 187 (1932). See also *ibid.*, **218**, 269 (1933); **249**, 111, 126 (1937).

<sup>25</sup> Rasmussen, K. E. *Compt. Rend. Trav. Lab. Carlsberg*, **20**, 11, 2 (1920); *ibid.*, **21**, 11, 2 (1921).

### Cosubstrates in Proteolysis

It is not improbable that the structure of clupein, as proposed by Waldschmidt-Leitz is approximately correct. Serious doubts are raised, however, as to the significance of the products isolated from proteins by enzymatic hydrolysis, owing to some recent work of fundamental importance by Bergmann and his coworkers. It was first shown<sup>28</sup> that typical proteolytic enzymes can readily effect not only hydrolysis but also synthesis of peptide bonds. Thus, in the presence of activated papain, benzoyl-*L*-leucine was found to react with *L*-leucine anilide, to form benzoyl-*L*-leucyl-*L*-leucine anilide; and many other similar syntheses could be achieved. This led to a further and even more important discovery; it was found that a peptide, not split by a given enzyme when alone, might be readily split in the presence of another peptide which itself could be recovered unchanged at the end of the reaction. For example, papain does not split glycyl-leucine, neither does it split acetyl phenylalanyl glycine. In the presence of both substrates, however, the enzyme breaks down the glycyl leucine<sup>29</sup> to glycine and leucine; acetyl phenylalanyl glycine is found unchanged at the end of the reaction. The interpretation given by Bergmann is as follows: a small degree of synthesis first occurs, and leucine is then split off by the enzyme from the synthesized peptide.



The enzyme then acts further to split off the terminal glycine residue, regenerating the acetyl phenylalanyl glycine, which thus acts as a "co-substrate," enabling papain to split glycyl leucine, although it is not itself split in the process.

The application of such experiments to protein hydrolysates is obvious. The process of degradation of a protein by an enzyme can no longer be looked upon as merely a progressive hydrolytic splitting to simpler and simpler molecules. Two peptides produced by hydrolysis may interact to give synthetic combinations; and such synthetic peptides may then undergo secondary breakdown to products quite different from those that would have been produced by the enzymes acting on either of the original peptides alone. The synthetic action of the enzyme may rearrange the amino acid residues into an order quite different from that existing in the original protein before hydrolysis. Therefore until the rules governing the possi-

can be drawn from enzymatic degradation studies concerning the grouping of amino acid residues in a protein molecule.

### The Hypothesis of Bergmann and Niemann

On the basis of analytical data, Bergmann and Niemann<sup>30, 31</sup> concluded that certain simple numerical relations held between the amino acids in any protein. Their hypothesis is that the total number of amino acid residues in a protein molecule is expressible by the formula  $2^m \times 3^n$ , where  $m$  and  $n$  are integers ( $\neq$  zero); and the number of residues of any individual amino acid is  $2^{m'} \times 3^{n'}$  where  $m'$  and  $n'$  are integers or zero. As one typical example of their calculations we may take silk fibroin (Table 2), which has an amino acid composition of remarkable simplicity<sup>32</sup>. Aside from the observed analytical values given in the table, an estimate of the mean residue weight of all amino acids formed on hydrolysis is necessary to

TABLE 2. RATIO OF AMINO ACIDS IN SILK FIBROIN AFTER HYDROLYSIS ACCORDING TO BERGMANN AND NIEMANN

Amino acid	Weight (per cent) (1)	Molecular weight (2)	Gram molecules per 100 grams protein		Reciprocal fraction of total number of residues (frequency)	
			Found (3)	Calculated (4)	Ratio (5)	(6)
Glycine.....	43.8	75	0.5840	(0.5840)	1,296	2
Alanine.....	26.4	89	0.2966	0.2920	648	4
Tyrosine.....	13.2	181	0.0729	0.0730	162	16
Arginine.....	0.95	174	0.0055	0.0054	12	216
Lysine.....	0.25	146	0.0017	0.0018	4	648
Histidine.....	0.07	155	0.00045	0.00045	1	2,592

The basis of the calculated values is the observed figure for glycine, 0.5840 in Column 3.  
Values for lysine and histidine from H. B. Vickery and R. J. Block, *J. Biol. Chem.*, 93, 105, (1931); other values from M. Bergmann and C. Niemann, *J. Biol. Chem.*, 122, 577 (1938).

calculate the figures given in the last two columns of the table. This mean residue weight was estimated as 84 for silk fibroin: a figure involving little doubt in this case, since most of the amino acids formed on hydrolysis have been determined.

If the analytical figure given for histidine is accepted, it follows that silk fibroin contains  $2592 = 2^5 \times 3^4$  amino acid residues per molecule, or some integral multiple of this number. If it is assumed that the histidine arises from an impurity which does not really belong to the silk fibroin molecule, and this figure is omitted, the minimum number is halved, to give 1296. In either case, the number of residues of each amino acid present (Table 2, Column 5) is expressible by the formula  $2^{m'} \times 3^{n'}$ . The reciprocal fraction of the total number of residues contributed by each amino acid (Table 2, Column 6) has been called by Bergmann and Nie-

conception that each amino acid occurs in the peptide chain at regular intervals; glycine forming every alternate residue, alanine every fourth, tyrosine every sixteenth, and so on. This, of course, does not necessarily follow even if it is true that the formula  $2^m \times 3^n$  holds for each type of residue; but the long periodic spacings revealed by x-ray diffraction data on proteins (Chapter 14) render some hypothesis regarding periodicity very attractive.

A relatively simple molecule, for which the formula of Bergmann and Niemann appears to hold, is secretin, the composition of which has been studied carefully by Ågren<sup>33</sup>. The total nitrogen content ( $14.53 \pm 0.18\%$ ) and sulfur content ( $0.675 \pm 0.005\%$ ) have been accurately determined, and the molecular weight has been estimated by ultracentrifugal measurement

TABLE 3. COMPOSITION OF SECRETIN\*

Amino acid	Gm moles/100 gm secretin		No. of residues per mole secretin (3)	Reciprocal fraction of total number of residues (frequency) (4)
	calculated (1)	found (2)		
Lysyl.....	0.063	0.062	3	12
Arginyl.....	0.042	0.039	2	18
Prolyl.....	0.042	0.042	2	18
Histidyl.....	0.021	0.026	1	36
Glutaminyl.....	0.021	0.020	1	36
Asparaginyl.....	0.021	0.021	1	36
Methionyl†.....	(0.021)	0.021	1	36

Values in Column 2 are from G. Ågren, *Skand. Arch. Physiol.*, 70, 10 (1934); *J. Physiol.*, 94, 553 (1938-39), on secretin chloride dried in vacuo over  $P_2O_5$  at  $50^\circ$ . All analytical figures are corrected for ash content of secretin.

\* From C. Niemann, *Proc. Nat. Acad. Sci.*, 25, 267 (1939).

† The basic figure (0.021) is that of the sulfur content, provisionally taken as methionyl sulfur, since no labile sulfur was found in secretin.

The glutaminyl and asparaginyl values are taken as the mean of one-half of the amide nitrogen as determined by the Van Slyke method.

to be near 5,000. The content of basic amino acids and of proline has also been determined. There is no labile sulfur; the sulfur present is therefore due presumably to methionine. Approximately one mole of aspartic and one of glutamic acid were isolated, per mole of secretin; also two molecules of amide nitrogen. Utilizing these figures, Niemann<sup>34</sup> has calculated that there are  $36 = 3^2 \times 2^2$  amino acid residues per mole of secretin, and the numbers of the known residues are as follows (Table 3).

Some residues of aspartic or glutamic acid, or both, may be present in addition to those isolated; for Ågren (33) has determined the isoelectric point of secretin as 7.5, and at this pH the molecule must carry five positive charges (three from lysyl and two from arginyl residues). Five negative charges must therefore be present to make the molecule isoelectric, and these would most probably be carboxylic groups arising from aspartic or

glutamic acid, with the terminal carboxyls not bound in amide linkage. The figures given in Table 3, for all residues yet isolated, are in accord with the hypothesis of Bergmann and Niemann.

This hypothesis has also been applied to the analytical data on blood fibrin, cattle hemoglobin, egg albumin, gelatin<sup>30, 31</sup>, insulin<sup>35</sup> and other proteins. In seeking further tests of its validity, two fundamental conditions must be met: (1) The protein to be analyzed must be a pure chemical individual, and it is often difficult to prove that this condition is satisfied. In any case, the protein to be studied must satisfy as many different criteria of purity as possible. Apart from analytical evidence, ultracentrifugation, electrophoresis and solubility measurements provide particularly important criteria<sup>36</sup>. All these are discussed in subsequent chapters. (2) The methods of analysis employed must be reliable within two or three per cent, and should be better if possible. Most methods used in the past for estimating the amino acid content of proteins have not met this standard, though recent important advances in this field, discussed in Chapter 15, give hope that in the future this standard may well be met. Every advance in the analytical and preparative chemistry of the proteins may be expected to advance also their structural chemistry.

<sup>35</sup> du Vigneaud, V., *Cold Spring Harbor Symp. Quant. Biol.*, 6, 275 (1938).

<sup>36</sup> Pirie, N. W., *Biological Reviews (Camb. Phil. Soc.)*, 15, 377 (1940).

## Chapter 14

### X-Ray Diffraction Studies and Protein Structure

BY JOHN T. EDSALL

#### Interatomic Distances in Amino Acids and Diketopiperazine

X-ray diffraction studies have already yielded valuable information concerning the nature of protein fibers and crystals, and they promise to yield far more in the future. Since it has been shown in the preceding chapter, however, that the amino acids in proteins are bound together by peptide linkages, we may first consider the nature of spatial arrangements to be expected in a simple extended polypeptide chain. As yet, the structure of no peptide has been determined by x-ray diffraction methods. Valuable evidence is, however, available from the very exact studies of Corey and his co-workers on glycine<sup>1</sup>, alanine<sup>2</sup> and diketopiperazine<sup>3</sup>. (The earlier studies of Bernal, and of Hengenstenberg and Lenel, on amino acids, are discussed in Chapter 8).

The structure of the diketopiperazine molecule is given by the diagram in Fig. 1. It is a nearly plane hexagon, with the angles between all bonds in the ring equal to  $120 \pm 3^\circ$ . Because of resonance in the peptide linkage,

between the structures  $\text{O}=\text{C}-\text{NH}-$  and  $\text{O}^--\text{C}^+=\text{NH}-$ , the C—O (1.25) and C—N (1.33) distances in this group are nearer to those characteristic of double than of single bonds. The N—CH<sub>2</sub> distance (1.41) is much shorter than the C—N single-bond distance (1.47) found in several other compounds, such as CH<sub>3</sub>NO<sub>2</sub>, CH<sub>3</sub>N<sub>3</sub>, CH<sub>3</sub>NC and N(CH<sub>3</sub>)<sub>3</sub><sup>4</sup>, but agrees well with that found in glycine (see below). The C—C distance (1.47) is also much shorter than the normal single-bond distance (1.54 Å).

"In crystals of diketopiperazine the molecules are linked together by hydrogen bonds between their respective oxygen atoms and =NH groups to form flat continuous chains throughout the structure." (Reference 4, p. 232.) (Fig. 2). Such hydrogen bonds may be important in the structure of polypeptides, and of protein crystals and fibers, since they offer a possible explanation for the cohesive forces which hold adjacent peptide

The form of the glycine molecule<sup>1</sup>, and the interatomic distances in it, are shown in Fig. 3. The  $\alpha$ -carbon atom lies in the plane of the  $\text{COO}^-$  group, while the nitrogen atom is slightly ( $0.268 \text{ \AA}$ ) above this plane. The interatomic distances (probable error  $\pm 0.02 \text{ \AA}$ ) and bond angles are entirely consistent with those found in other molecules with similar bonds,

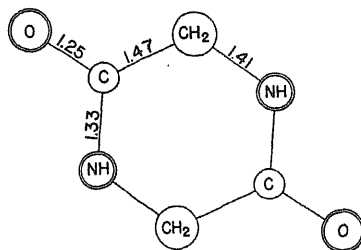


Figure 1. A diagram of the molecule of diketopiperazine. Figures give interatomic distances in Angstrom units. Angles between all bonds are close to  $120^\circ$ . (From R. B. Corey, *Chem. Rev.*, 26, 227 (1940), Fig. 1, p. 230).

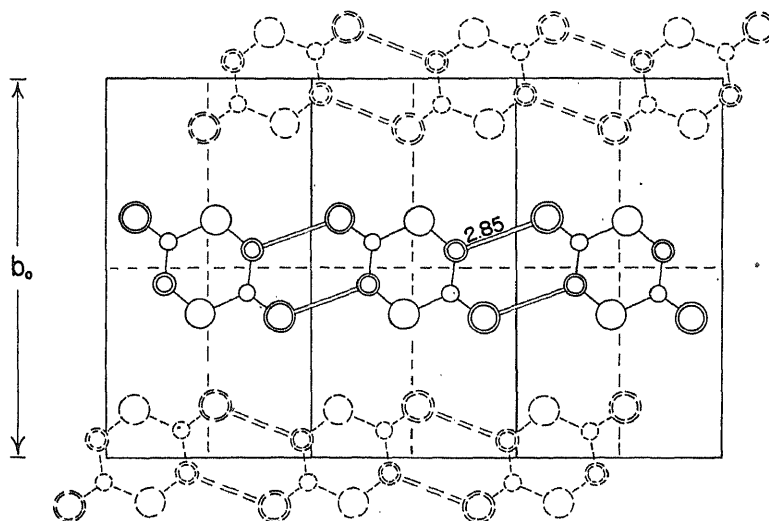


Figure 2. View perpendicular to the (101) plane of the diketopiperazine crystal, showing chains of molecules held together by hydrogen bonds (double lines)  $2.85 \text{ \AA}$  in length between oxygen atoms and  $=\text{NH}$  groups. (From R. B. Corey, *Chem. Rev.*, 26, 227 (1940), Fig. 2, p. 232.)

except for the  $\text{H}_2\text{C}-\text{N}$  distance ( $1.39 \text{ \AA}$ ) which, like that in diketopiperazine, is lower than that found in many other compounds. In the original communication the structure of the glycine crystal is given in great detail



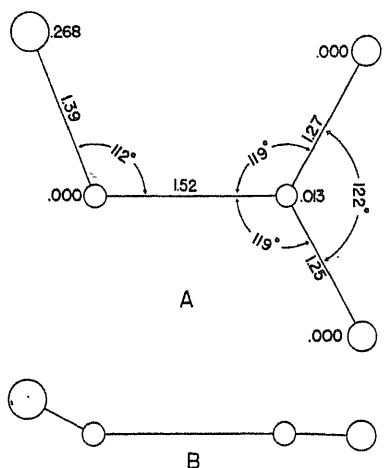
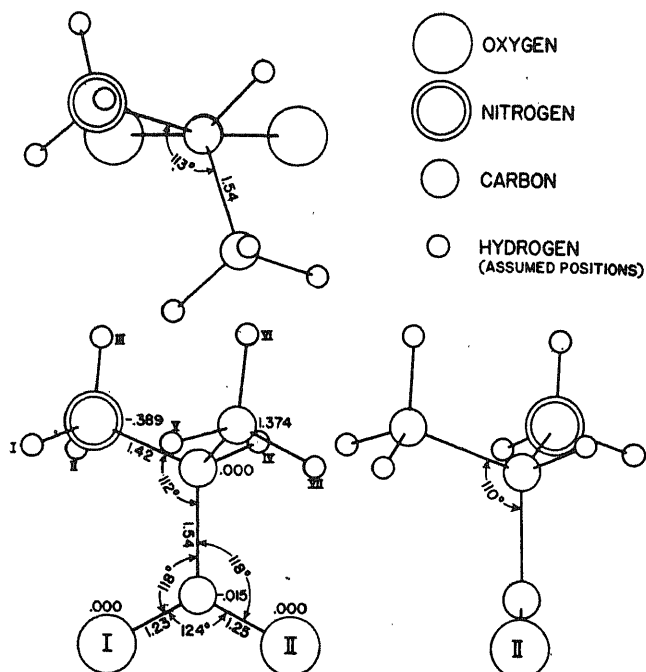


Figure 3. The molecule of glycine viewed (A) perpendicular to a plane containing the two oxygen and the  $\alpha$ -carbon atoms, and (B) parallel to this plane. (From R. B. Corey, *Chem. Rev.*, 26, 227 (1940), see Fig. 3, p. 233.)



The same is true of the very careful and detailed study of the crystal structure of *dl*-alanine by Levy and Corey<sup>2</sup>. Here again we shall give only the configuration of the alanine molecule (Fig. 4). This is entirely in accord with the findings on glycine and other simple molecules, as regards both interatomic distances and bond angles; the C—CH<sub>3</sub> distance shows

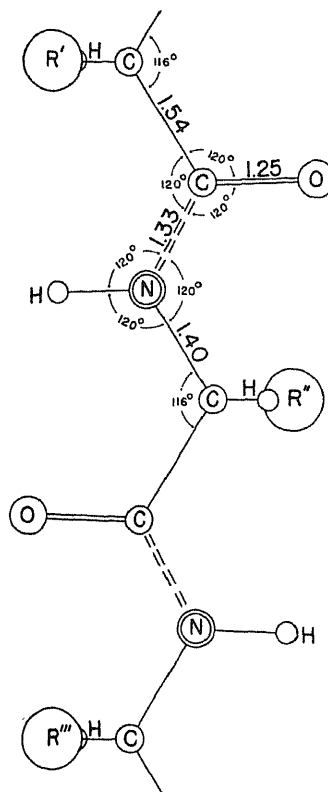


Figure 5. A diagrammatic representation of a fully extended polypeptide chain, based upon the interatomic distances and bond angles found in crystals of diketopiperazine and glycine. The double-dashed lines represent resonating bonds (1.33 Å) between the ketocarbon and nitrogen atoms. (From R. B. Corey, *Chem. Rev.*, 26, 227 (1940), Fig. 4, p. 233.)

the usual value of 1.54 Å; the C—N distance is 1.42 Å, a little longer than in glycine but identical within the limits of experimental error. Other dimensions are shown clearly in the figure.

#### Inferences Concerning the Structure of Polypeptide Chains

values chosen for the bond distances in this chain are obvious in terms of the data shown in Figs. 1, 3 and 4. The bond angles about the carbonyl carbon and the nitrogen atoms are assumed to be  $120^\circ$  as in diketopiperazine. Around the  $\alpha$ -carbon atoms the bonds are assumed to be arranged in nearly tetrahedral fashion, except that the C—C—N bond angle is taken as  $116^\circ$ , a mean of the values found in glycine ( $112^\circ$ ) and diketopiperazine ( $120^\circ$ ). The fully extended chain is coplanar, with the C=O oxygen atoms and the N—H hydrogen atoms in the plane of the chain. The distance along the chain corresponding to one amino acid residue is 3.67 Å. Since the spatial configuration of the groups around the  $\alpha$ -carbon atom on all amino acid residues found in proteins is the same, the side chains  $R'$ ,  $R''$ ,  $R'''$  . . . attached to the  $\alpha$ -carbon atoms point alternately, first ( $R'$ ,  $R'''$ ) to the left and upward from the plane of the chain, next ( $R''$ ) to the right and downward from this plane.

We may now consider the relation of this model to actual proteins which have been studied by x-ray diffraction methods. These fall into two great classes. The fibrous proteins include such substances as silk fibroin, keratin, myosin, and collagen. These, as their name implies, are found in nature as essential components of long, thin fibrous structures (silk fiber, wool or hair, muscle, tendon). The so-called "globular" proteins include practically all the truly crystalline proteins so far examined, such as egg albumin, crystalline serum albumin, the hemoglobins, the hemocyanins, pepsin, trypsin, and the other crystalline enzymes thus far described. Many of these proteins, perhaps most, are not truly globular in the sense of being spherical<sup>5</sup>, but in contrast to the extremely elongated fibrous proteins they deviate relatively little from the spherical shape. It is among the fibrous proteins, if anywhere, that we may expect to find structures resembling the elongated polypeptide chain pictured in Fig. 5. This class of proteins may therefore be considered first.

### The Fibrous Proteins

Many sorts of protein fibers from different tissues have been examined, but all appear to belong to one of two groups: (1) the keratin-myosin group; (2) the collagen group. Within the former group, the first protein studied in detail by x-ray methods was silk fibroin<sup>6</sup>. The chemistry of this protein is of remarkable simplicity<sup>7</sup>; one-half of all the amino acid residues are glycine, one-quarter alanine, one-sixteenth tyrosine. Definite but small amounts of lysine and arginine are present, and an extremely small quantity of histidine.

The x-ray photographs<sup>6</sup> of silk fibers show many well defined spots, and have a considerable resemblance to those obtained on simple crystalline substances. In particular, there is a well marked periodicity of  $7.0 \pm .2 \text{ \AA}$  along the fiber axis. This fits extremely well with the hypothesis of a repeating unit, two amino acid residues in length, along this axis. The distance along the chain, per residue, would then be  $3.5 \text{ \AA}$ , only slightly below that deduced from the model shown in Fig. 5. The combination of the x-ray and chemical evidence here leaves little doubt that the silk fiber is made up of extended polypeptide chains. That the chains are almost fully extended is an assumption in accord with the mechanical properties of the silk fiber, which has only a small range of reversible elasticity, and always becomes permanently elongated when stretched more than a few per cent. The stretching brings about no fundamental change in the x-ray pattern—in marked contrast to the behavior of keratin on stretching, which is discussed below—and presumably involves only a displacement of the different polypeptide chains relative to one another, along the line of the fiber axis.

A number of spacings in the equatorial plane (perpendicular to the fiber axis) have been reported for silk fibroin, the most prominent being at  $4.3$  and  $4.6 \text{ \AA}$ . Several attempts have been made to define a unit cell for the fiber on the basis of all these data. Meyer and Mark<sup>6</sup> for example showed that the x-ray data could be fitted by a monoclinic unit cell ( $a = 9.68 \text{ \AA}$ ,  $b = 7.0 \text{ \AA}$ ,  $c = 8.80 \text{ \AA}$ ,  $\alpha = 75^\circ 50'$ ) containing four parallel glycylalanyl residues, glycylalanyl being regarded as the repeating unit<sup>8</sup>. Unfortunately such a cell would not accommodate the tyrosyl residues, which are invariably found as part of the fibroin structure, and various special and uncertain hypotheses have been proposed to meet this difficulty; in particular it has been suggested that the tyrosine is part of an amorphous structure within the fiber, and is distinct from the chain giving the true crystalline pattern<sup>9</sup>. The situation needs further clarification; but the correlation of the  $7 \text{ \AA}$  spacing along the fiber axis with an extended peptide chain appears in any case to be well founded.

The fibrous protein keratin, as found in hair, horn, or wool, also gives well defined fiber diagrams, although these are generally not so sharp or so rich in detail as those of silk fibroin. The principal spacing in the direction of the fiber axis is not  $7 \text{ \AA}$ , but  $5.1 \text{ \AA}$ , and the other details of the photographs are quite different. Keratin also differs radically from silk fibroin in being reversibly extensible, up to about double its original length, the stretching occurring with relative ease when the hair is soaked in hot water or in dilute solutions of alkalis. The x-ray diagram given by

keratin in the fully extended hair is quite different from that of the unstretched hair, showing a much greater resemblance to that of silk fibroin. In particular, there is a well marked spacing, along the fiber axis, of approximately  $3.4 \text{ \AA}$ <sup>10</sup>. This is significantly lower than the  $3.67 \text{ \AA}$  calculated from Fig. 5 for a single residue in a fully extended polypeptide chain, but it is entirely compatible with an extended but slightly distorted chain, perhaps differing from the fully extended chain by rotations around the C—C bonds (see <sup>4</sup>). Astbury has called this extended form of keratin the  $\beta$  form, the unstretched form found in natural hair being known as  $\alpha$ -keratin<sup>11</sup>.

The  $\beta$ -keratin fiber diagram shows two prominent equatorial spacings, at 10 and at  $4.65 \text{ \AA}$ . These spacings are not only at right angles to the axis of the fiber; they are also at right angles to each other. In  $\beta$ -keratin, as ordinarily prepared, this is not apparent; but Astbury and Sisson<sup>12</sup> succeeded, by lateral compression of keratin (horn) in steam, in bringing about a reorientation of the fibrous molecules so that the  $9.8 \text{ \AA}$  spacing stands normal to the plane of flattening, and the  $4.65 \text{ \AA}$  spacing lies in this plane. The horn must be treated with steam during the process of compression in order to make the structure more labile.

A highly probable interpretation of these spacings has been given by Astbury<sup>13</sup>. We may picture an assembly of polypeptide chains, each chain similar to that in Fig. 5, lying side by side in the plane of the paper, the distance between the axes of adjacent parallel chains being  $4.65 \text{ \AA}$ . The side chains  $R'$ ,  $R''$ ,  $R''' \dots$  project above and below this plane. Other sets of parallel peptide chains lie above and below the plane of the paper, being separated due to the side chains by an interval of  $10 \text{ \AA}$ . The main chains, separated by the "backbone" spacing of  $4.65 \text{ \AA}$ , may well be held together by hydrogen bonds ( $\text{C}=\text{O} \dots \text{H}-\text{N}$ ) between the CO and NH groups in adjoining peptide chains<sup>14</sup>. A diagram of this structure is given in Fig. 6. The association of the  $9.8 \text{ \AA}$  spacing with the side chains is made more probable by the observation<sup>15</sup> that this spacing alone is altered when keratin and other related proteins absorb water. It changes from about  $9.8 \text{ \AA}$ , in a dry protein, to as much as  $10.8$  or more in the same protein when wet. The entering water would naturally tend to fit in most readily among and between the side chains in this structure, leaving the  $3.4$  and  $4.65 \text{ \AA}$  spacings practically unaltered.

<sup>10</sup> Astbury, W. T., and Street, A., *Phil. Trans. Roy. Soc., London*, **A230**, 75 (1931); Astbury, W. T., and Woods, H. J., *ibid.*, **A232**, 333 (1933-34).

<sup>11</sup> Concerning the exact value of the  $3.4 \text{ \AA}$  spacing, Astbury and Woods observed that it was "Owing to an inherent lack of definition and to the fact that the spacing is not constant throughout the length of biological subjects can rarely be more than an approximate value." The spacing of the  $(020)$  reflection along the fiber axis of  $\beta$ -keratin is, like that of  $\alpha$ -keratin, slightly variable according to previous treatment and the state of tension. The mean value of the spacing of the  $(020)$  arc appears to be  $3.38 \text{ \AA}$ .

If a fiber of  $\beta$ -keratin is released soon after stretching, it contracts again to the  $\alpha$ -form. Since the transformation is reversible, the peptide chain is presumably unruptured in both forms. Since in  $\beta$ -keratin the chain is nearly fully extended, in the  $\alpha$ -form—only half as long—it must be somehow folded. Furthermore, the folding must take place in the plane of the chains—the plane of the paper in Fig. 5—for the side chain spacing remains practically unaltered by the change between the  $\beta$  and  $\alpha$ -form. The 5.1 Å spacing may be associated with a unit of three amino acid residues in the peptide chain; since 5.1 Å in the  $\alpha$ -form is equivalent very

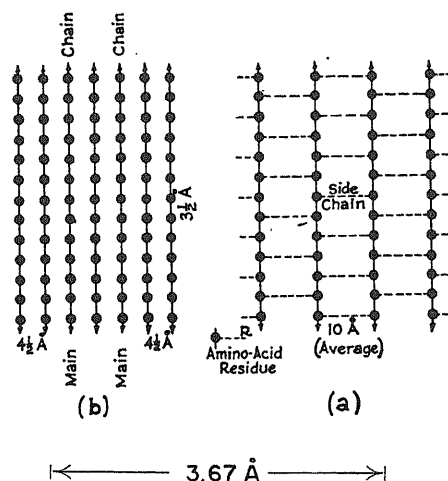


Figure 6. Diagrammatic scheme for the structure of  $\beta$ -keratin (stretched wool, hair, etc.). (a) View perpendicular to an extended polypeptide grid. (b) View parallel to a series of grids fitted face to face to form a sub-microscopic  $\beta$ -crystallite. The  $\alpha$ -configuration may be derived from (a) by folding the main-chains in planes roughly perpendicular to that of the paper. (After W. T. Astbury and R. Lomax, *J. Chem. Soc.*, 1935, p. 847.)

nearly to 10.2 Å in the  $\beta$  form; and  $10.2 = 3 \times 3.4$ , the length of the principal axis spacing in  $\beta$ -keratin. At present, the nature of the folds in  $\alpha$ -keratin cannot be more precisely defined. A specific model for  $\alpha$ -keratin earlier suggested<sup>16</sup> has been shown by Neurath<sup>17</sup> to involve closer approach between the atoms of adjoining side chains than is possible in the light of modern crystal structure data; and Astbury<sup>14</sup> has therefore abandoned his earlier scheme. A model for the folding involved in the  $\alpha$ - $\beta$  transformation has very recently been proposed by W. T. Astbury and F. O. Bell<sup>17a</sup>. The number of possible types of folding which would fit all the

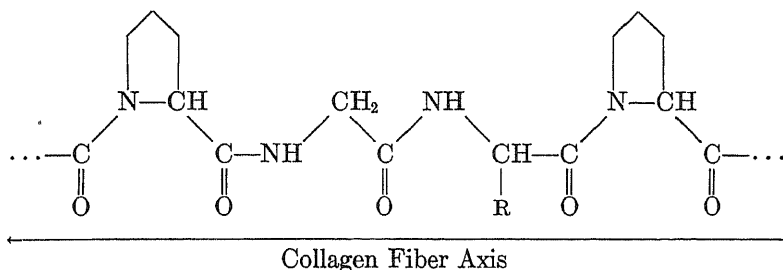
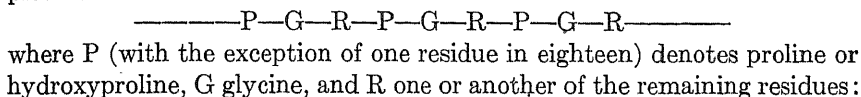
present experimental data is probably very great, and much further work is required to determine the nature of the actual pattern.

A keratin fiber stretched in steam or dilute alkali, and then immediately released, will often contract to about two-thirds of its initial unstretched length, a phenomenon termed "supercontraction" by Astbury and Woods (10). The effect of the steam or alkali is to loosen the linkages between adjacent side chains, and thereby render it easier for the peptide chains to fold or unfold from one configuration to another. The supercontracted form of keratin must obviously involve an even more extensive folding of the peptide chain than that found in  $\alpha$ -keratin.

The principal globulin of muscle, myosin, can be drawn into fibers or dried in films and stretched. These fibers or films give an x-ray diagram strikingly similar to that given by muscle<sup>18, 19</sup>, which in the relaxed state gives a fiber diagram remarkably similar to that of keratin. There is little doubt that the optical and mechanical properties of muscle are determined primarily by bundles of myosin molecules lying parallel to the axis of the fiber. It is therefore of great importance that Astbury and Dickinson<sup>19</sup> have shown that myosin can be obtained in  $\alpha$ ,  $\beta$  and supercontracted forms, closely comparable to those of keratin, and with spacings identical within the limits of experimental error. They suggest that, since the myosin in relaxed muscle is in the  $\alpha$ -form, the process of muscular contraction involves a transformation from  $\alpha$  to supercontracted myosin. Myosin is, of course, very different from keratin in its chemical composition, notably in its much lower cystine content, but there is evidently a common underlying pattern of molecular architecture in the two compounds, based on the general pattern of the peptide chain and the manner of its folding. The true resemblance, however, is not between myosin and natural keratin, but between myosin and keratin which has been made labile by steam or alkali, so that it has become capable of supercontraction; the links between side chains in natural keratin are far more tenacious than in myosin. This greater lability of myosin than of keratin corresponds to its biological function, and may be correlated with its much lower cystine content<sup>20</sup>.

Collagen, as found in connective tissue, tendon, cartilage, and other tissues, represents a fibrous protein built apparently on a different plan from the silk fibroin-keratin-myosin group. Chemically, collagen fibers are notable for their high content of proline and hydroxyproline, amounting to a little less than one-third of all the residues present; glycine is generally present to about another third; the other residues are made up by a variety

of amino acids. The principal spacing along the fiber axis is about 2.86 Å, and this appears actually to correspond to the average length along the chain per amino acid residue. There is a "backbone" spacing of about 4.4 Å (very diffuse), and a side-chain spacing of about 11.5 Å (in *wet* gelatin). The point most requiring explanation is that the 2.86 Å spacing is so much shorter than the 3.4 Å spacing in  $\beta$ -keratin, although collagen fibers are practically inextensible. Several models have been proposed to explain this, by Astbury and others. In the latest<sup>21</sup>, the peptide chain is pictured as of the form:



Astbury's paper gives details of the spatial configuration of the model. The total length along the chain for three residues works out as 8.55 Å, or 2.85 per residue. This picture is certainly attractive, and skilfully takes account of the high proline and glycine content of collagen. Final tests of its validity remain to be obtained.

**The "Globular" Proteins.** These, the true protein crystals, differ from the fibrous proteins in giving, in the undenatured state, x-ray diffraction patterns of extraordinary complexity and very sharp detail, with a very great number of different spacings. Such photographs could not, however, be obtained until the necessary conditions were found to prevent alteration of the crystal. Bernal and Crowfoot<sup>22</sup> found the essential condition to be that the crystals be surrounded by mother liquor, since even moderate drying of most protein crystals produces marked alterations in the x-ray diagrams. The usual technique is to draw up a single crystal, surrounded by its mother liquor, into a thin-walled capillary tube, which is then sealed. The tube and its contents can then be mounted, and exposed to the x-ray beam. Bernal, Crowfoot, Fankuchen, Riley and others have now examined pepsin, chymotrypsin, horse methemoglobin, lactoglobulin (two crystal forms), insulin, serum albumin, ribonuclease and a



TABLE 1. X-RAY DATA ON CRYSTALLINE PROTEINS <sup>®</sup>  
(Prepared by I. Fankuchen)

	Best Non-x-ray Molecular Weights	Condi- tion Wet or Dry	"a", Å (10 <sup>-8</sup> cm)	"b", Å	"c", Å	$\beta$ in de- grees	$c \times$ $\sin \beta$	Volume in Å <sup>3</sup>	n	Volume per molecule (V)	Densi- ty ( $\rho$ )	Molecular Weight = $V\rho/1.65n$	Molec- ular Weight Cor- rected for Resid- ual Water	Space Group	Small- est Ob- served Spac- ing
iclease orthombic)	13,000 to 15,000	wet dry	36.6	40.5	52.3	90	52.3	77,300	4	19,300	1.341	15,700	13,700	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	2
iclease oclinic)	13,000 to 15,000	wet dry	30.8	38.5	53.5	107	51	60,000	2	30,000				P2 <sub>1</sub>	2
*	35,100 to 40,900	wet dry	28.7	29.3	45.2	100	44.5	37,400	2	18,700				P2 <sub>1</sub>	3.7
			144	83	34	90	34	404,000	6	67,000	1.28	52,400		R3	2.4
lobulin lar form)	37,900 to 41,800	wet dry	130	74.8	30.9	90	30.9	298,000	6	50,000	1.315	39,500	37,400	R3	7
			67.5	67.5	154	90	154	702,000	8	88,000	1.257	67,000		P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	2.4
lobulin lle form)	37,900 to 41,800	wet dry	60	63	110	90	110	416,000	8	52,000	1.27	40,000		P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	20
			67.5	67.5	133.5	90	133.5	608,000	8	76,000				P4 <sub>2</sub> 2 <sub>1</sub>	
notrypsin	27,000	wet dry	56	56	130	90	130	408,000	8	51,000	1.30	40,100			20
			69.5	69.5	97.5	90	97.5	471,000	8	58,900				P4 <sub>2</sub> 2 <sub>1</sub> P4 <sub>2</sub> 2 <sub>1</sub>	2.5 10
trypsin	41,000	wet dry	63.0	63.0	74.5	90	74.5	298,000	8	37,200	1.33	30,100			
			49.6	67.8	66.5	102	65	219,000	2	109,000	1.277	84,500	54,000	P2 <sub>1</sub> P2 <sub>1</sub>	2 5
	35,500- to 39,200	wet dry	45	62.5	57.5	112	53.5	151,000	2	75,500	1.31	60,000			
			116	67	461	90	461	3,580,000	54	66,500	1.32	53,000			

methemo- in	66,700	wet dry	110 102	63.8 51	54.2 47	112 130	50.2 36	352,000 188,000	2	176,000 94,000	1.242 1.27	132,000 72,000	66,700	C2 C2	2 13
serum min	70,000 73,000	wet dry	96.7 74.5		145 130	120 120		1,170,000 610,000	6	195,000 102,000	1.27 1.34	150,000 82,800		H H	4 20
co seed ulin	(300,000)	wet dry												F	
in	294,000	wet dry	123 149	123 86	123 208	90 90	123 208	1,860,000 2,670,000	4	465,000 445,000	1.287 1.31?	362,000 350,000	322,000 305,800		
stunt s	7,600,000 10,600,000	wet dry	394 318	394 318	394 318	90 90	394 318	61,000,000 32,000,000	230 216	500,000 16,000,000	1.286 1.35	24,000,000 13,000,000		R3 I	

values given here for insulin are referred to an hexagonal unit cell. The described on the basis of a rhombohedral lattice gives  $a = 44.4$ ,  $\alpha = n = 1$ , for dry insulin; and  $a = 49.4$ ,  $\alpha = 114^\circ 16'$ , and  $n = 1$ , for wet insulin kuchen, I., *J. Gen. Physiol.*, **24**, 315 (1940-41).

- foot, D., *Proc. Roy. Soc. London*, **A 164**, 580 (1938).  
 foot, D., *Nature*, **135**, 691 (1935).  
 foot, D., and Riley, D., *Nature*, **144**, 1011 (1939).  
 foot, D., and Riley, D., *Nature*, **141**, 521 (1938).  
 foot, D., and Riley, D., Unpublished data (1940).  
 al, J. D., Fankuchen, I., and Perutz, M., *Nature*, **141**, 523 (1938).  
 itz, M. F., Ph.D., Thesis, University of Cambridge (1939).
- (10) Bernal, J. D., and Crowfoot, D., *Nature*, **133**, 794 (1934).  
 (11) Crowfoot, D., and Fankuchen, I., *Nature*, **141**, 523 (1938).  
 (12) Astbury, W. T., Dickinson, O., and Bailey, K., *Biochem. J.*, **29**, 2351 (1935).  
 (13) Astbury, W. T., and Bell, F. O., unpublished.  
 (14) Astbury, W. T., and Bell, F. O., *Gen. Physiol.*, **17**, (1), 90 (1939).  
 (15) Bernal, J. D., Fankuchen, I., and Riley, D., *Nature*, **142**, 1075 (1938).  
 (16) Bernal, J. D., *J. Gen. Physiol.*, **24**, 293 (1940-41).  
 (17) Bernal, J. D., *Proc. Roy. Soc. London*, **B 127**, 1 (1939).  
 (18) Svedberg, M., and Northrop, J. H., *J. Gen. Physiol.*, **18**, 433 (1935).  
 (19) McFarlane, A. S., and Kekwick, R. A., *Biochem. J.*, **32**, 1607 (1938).  
 (20) Lauffer, M. A., and Stanley, W. M., *J. Biol. Chem.*, **135**, 463 (1940).  
 (21) Crowfoot, D., *Chem. Rev.*, **28**, 215 (1941).

*Chymotrypsin*. Bernal, Fankuchen and Perutz (8), Perutz (9). (In the original communication (8)  $n$  was taken as 4. It is now suggested by Perutz that  $n = 2$ . This gives a molecular weight in better agreement with the data.) Crystals prepared by Northrop.

*Hemoglobin*. Bernal, Fankuchen and Perutz (8). Further intensive study is being done by Perutz (9). See also Crowfoot (21). Perutz, *Nature* **149**, 491 (1942).

*Pepsin*. Bernal and Crowfoot (10). These measurements were the first on single protein crystals and repetition of this work would be desirable.

## X-RAY DATA

- case (orthorhombic)*. I. Fankuchen (1).  
*case (monoclinic)*. I. Fankuchen (2).  
 1. Crowfoot (3, 4) (dry), and Crowfoot and Riley (5) et).  
 2. Crowfoot and Riley (6). Further work is being carried on by these workers (private communication) including Fourier studies and x-ray studies of crystals in various states of hydration (7). See Crowfoot (21).  
*otrypsin*. prepared by Kunitz. X-ray studies by I. Fankuchen (1).

TABLE 1—Continued

NON-X-RAY MOLECULAR WEIGHT DATA	DENSITIES
<p><i>rum albumin</i>: prepared by T. L. McMeekin; x-ray es by I. Fankuchen (1).</p> <p><i>Seed Globulin</i>. Crowfoot and Fankuchen (11). Fairly dry crystals were used prepared by Vickery. Only extremely faint powder lines were observed. It d be desirable to have these measurements repeated an effort to obtain data on wet crystals would be h while.</p> <p>Astbury, Dickinson and Bailey (12) and Astbury Bell (13). Astbury and Bell refer to this unpublished in their paper in <i>Tabulae Biologicae</i> (14).</p> <p><i>Seed Virus</i>. Bernal, Fankuchen and Riley (15). rial furnished by Bawden and Pirie, and x-ray studies ernal, D., and Fankuchen, I., <i>J. Gen. Physiol.</i>, <b>25</b>, 147 (1941).</p>	<p><i>Ribonuclease</i>: I. Fankuchen (1).</p> <p><i>Insulin</i>: both needle-shaped and flat rhombohedra were studied by Crowfoot. These were shown to possess identical crystal structure. The needles were imperfect and gave density values a trifle lower than the rhombohedra. The value chosen was the highest observed and was obtained from the largest rhombohedral crystals. Residual water 5.35% in air-dried crystals was determined by drying under reduced pressure at 104°.</p> <p><i>Lactoglobulin</i>: Wet Tabular, 1.257 in sugar solution (Crowfoot, 6). Wet Needles, dissolved too rapidly to permit of any density determinations. Dry Crystals found to be 1.27 by immersion in O-dichlorobenzene and toluene (6). Crow- foot believes this value too low, due to occlusion of air and uses an assumed density of 1.31 (insulin) in computations of molecular weights.</p> <p><i>Chymotrypsin</i>: Wet. Determined by Perutz (8). Dry, value of 1.31 assumed (8).</p> <p><i>Hemoglobin</i>: Wet. Determined by Perutz (8). Dry, value of 1.26 assumed to correspond with value for serum albumin as given by Chick and Martin.</p> <p><i>Tobacco Seed Globulin</i>: measured in sodium phosphate buffer solution at pH 5.0. Residual water 10.4%. Determined by drying in vacuum at 100°.</p> <p><i>Excelsin</i>: Value of 1.31 assumed by I. Fankuchen to permit computation of molecular weight. Table in paper by Astbury and Bell gives only unit cell data and dry molecu- lar weight. Assumption is made that this value is cor- rected for residual water.</p> <p><i>Bushy Stunt Virus</i>: Wet crystals, 1.286, measured by immersion. Dry crystals, 1.35, from computations of McFarlane and Kekwick (19).</p>

tobacco seed globulin by this method. For each of these substances the crystal class, and the form of the unit cell, have been determined. In most cases they have studied the crystals both in the wet state and when dry (partly denatured). If the density,  $\rho$ , of the crystals is also known, a simple way of determining the molecular weight ( $M$ ) is available, for if the unit cell is of volume  $V$ , and contains  $n$  molecules, the mass of one molecule  $V\rho/n$ , and the molecular weight is  $N = 6.022 \times 10^{23}$  times as great; whence

$$M = \frac{V\rho}{n} \times (6.022 \times 10^{23}) \times 10^{-24} = \frac{V\rho}{1.66n}$$

if  $V$  is expressed in cubic Ångströms ( $1 \text{ Å}^3 = 10^{-24} \text{ cm}^3$ ).

" $n$ " is always a small integer, the value of which may generally be readily determined from other data which approximately fix the molecular weight of the protein. It is to be noted that the molecular weight calculated from this formula, for a hydrated crystal, is the molecular weight of the protein hydrate, not that of the anhydrous protein.

Unit cell dimensions, densities, and molecular weights of a number of proteins are given in Table 1. These data give information, not only concerning molecular size, but in some cases also concerning shape. Thus Perutz<sup>23</sup> has concluded from a study of hemoglobin, which crystallizes with a monoclinic unit cell containing two molecules, that the molecules cannot fit into the cell unless their shape does not differ widely from the spherical; probably the ratio of the long to the short axis cannot be greater than 2:1. Crowfoot and Riley<sup>24, 24a</sup> have drawn similar conclusions concerning the shape of the lactoglobulin molecule. In this case there are eight molecules in the unit cell; but given the dimensions of the cell and the size of the molecule, the possible type of packing in the cell severely limits the possible shapes which the molecule might assume, so that the axial ratio is almost certainly not greater than 2:1 or 3:1<sup>5</sup>.

These simple deductions can be drawn with the aid of only a very meager portion of the wealth of detail shown in the crystal photographs. A very great number of well defined spacings are revealed, some very long (over 100 Å), others as low as 2 or 3 Å. Hemoglobin, lactoglobulin, chymotrypsin and insulin crystals, in the wet state, all show very definite reflections corresponding to spacings of only 2 Å. "This", as Bernal, Fankuchen and Perutz<sup>25</sup> remark, "proves the complete internal regularity of the protein molecules down to atomic dimensions." In the above mentioned crystals, all these smaller spacings disappear from the photographs when the crystal is dried; the dry crystals show no spacings lower than about 8 Å (in hemoglobin none below 20 Å). Recently, however, Fankuchen<sup>26</sup> has

shown that dry crystals of the enzyme protein ribonuclease show definite reflections corresponding to spacings as low as 3 Å. Since the molecular weight of this protein is only about 13,000 (4 molecules per unit cell), further study of the crystals may yield results of very great interest.

**Virus Proteins.** Important results have been obtained by the study of tobacco mosaic virus and other viruses<sup>27, 28</sup>. Tobacco mosaic virus in a solution of medium concentration in water divides into two layers on standing<sup>28</sup>. The top layer (concentration 1.6 to 4%) is an isotropic liquid which becomes birefringent during flow, indicating the presence of elongated asymmetrical particles (see Chapter 21). The bottom layer, in equilibrium with it, is spontaneously doubly refractive and consists of regions in which the particles are parallel. The x-ray studies of Bernal and Fankuchen<sup>28</sup> have shown that the particles are also equidistant, and in cross-section are arranged in a simple hexagonal pattern (two-dimensional hexagonal close packing). The distances between the particles depend only on the concentration, showing that their distribution is homogeneous. Very concentrated preparations (above 30 per cent virus) have gel-like properties, and become stiffer with decreasing water content, but there are no abrupt transitions. The minimum distance between the particles in the wet gels is about 175 Å, but the hexagonal pattern is the same in these gels as in the most dilute bottom layer solutions studied by x-rays. Only the spacing of the pattern varies with concentration. On drying, further shrinkage occurs, and the air-dried gel has an interparticle distance of only 152 Å. "In these gels the arrangement of particles in the cross-section is so perfect that each specimen is a two-dimensional single crystal. The conclusion of this study is that the virus preparations consist of approximately cylindrical particles about 150 Å in diameter." (Reference 28, p. 163.) The length of the tobacco virus particles is at least ten times their diameter (see Chapter 21). The forces which maintain them equidistant and parallel in the gels are probably due to the ionic atmospheres surrounding them<sup>28a, 28b</sup>.

The foregoing discussion deals with the arrangement of the virus particles in solution relative to one another; it is based on x-ray measurements of spacings which vary systematically with concentration. Other spacings are found, however, which are quite independent of concentration and are determined only by the internal structure of the virus particle. These particles, even in solution, have an inner regularity like that of a crystal. "The structure seems to consist of sub-units of the dimensions of approxi-

<sup>27</sup> Winkler, R. W. G. and Carey, R. B. *J. Biol. Chem.* 116: 51 (1938).

mately 11 Å cube, fitted together in a hexagonal or pseudohexagonal lattice of dimensions— $a = 87$  Å,  $c = 68$  Å . . . the particle seems to be virtually unchanged by drying and must therefore contain little water" (reference 28, pp. 163–164). Although the spacings found markedly resemble those both of the fibrous and the globular proteins, tobacco mosaic virus exhibits definite differences from the proteins of either class.

Most other plant viruses so far studied are like tobacco mosaic virus in consisting of very elongated particles. The tomato bushy stunt virus<sup>28</sup>, however, crystallizes in the cubic system, and x-ray measurements suggest that the particles are not far from spherical. (See Table 1, also Chapter 19.)

### Tentative Inferences Concerning Protein Structure

All the "globular" proteins are readily hydrolyzed by proteolytic enzymes. Certainly, therefore, they contain polypeptide chains, although these chains must be built into a very systematic pattern of folds to explain the very highly ordered structure revealed by x-ray photographs of protein crystals. A partial breakdown of this order, with disappearance of the shorter spacings, occurs when the crystals are dried. On treatment with denaturing agents, such as urea, a further breakdown of this ordered structure occurs and the resulting products aggregate on coagulation into fibrous bundles which have been shown by x-ray study to be of the  $\beta$ -keratin type<sup>29</sup>. This work, largely on egg white and the seed globulins, indicates an intimate relation between the fibrous and the globular proteins. It implies that the globular proteins are potentially fibrous, and actually become fibrous in the last stages of denaturation. There are, of course, a very great number of intermediate stages between the highly organized protein crystal and the final product of its degeneration into a polypeptide chain of the  $\beta$ -keratin type.

There are grounds also for believing that the polypeptide chains themselves show a systematic pattern, a certain sequence of amino acid residues repeating at regular intervals. The existence of such regularities is indicated also by the analytical studies of Bergmann and Niemann, considered in Chapter 13, which suggest simple ratios between the frequencies of different amino acid residues in many proteins, both fibrous and globular. Such a sequence might arise, either from the juxtaposition of similar molecules in a crystal, or from the regular repetition of a certain pattern of residues over and over again in a long polypeptide chain of the fibrous type<sup>30</sup>. Either of these arrangements could give rise to very long spacings, and still other arrangements might be possible.

have pointed out that the feather keratin of the fowl shows spacings of 115 Å *normal to the fiber axis*. Since the work of Astbury indicates that the fundamental units in keratin fibers are parallel polypeptide chains, these chains must be arranged in such a fiber with a high degree of pattern and regularity relative to one another. The whole x-ray pattern of feather keratin, indeed, shows a richness of detail and a complexity of structure that rivals that of the crystalline globular proteins. Both types of protein certainly show a far higher degree of regularity in structure than can be observed in stretched fibers of denatured proteins. Both the "fibrous" and the "globular" proteins show evidence of a complex pattern of molecular and crystalline architecture whose details are yet to be unravelled.

### The Cyclol Hypothesis

A detailed hypothesis concerning protein structure has been put forward by D. M. Wrinch<sup>31</sup>. This rests initially upon the assumption of a lactim-lactem transformation similar to that at one time suggested by Astbury for  $\alpha$ -keratin<sup>16</sup>, involving two amino-acid residues, which are supposed to form a closed hexagon. In Wrinch's theory, however, these hexagons ("cyclols") are supposed to be closely joined and organized into an extended and highly systematic pattern. The simplest example of such a pattern is "cyclol 6" containing six amino-acid residues (Fig. 7). This is the prototype of many other more extended patterns which may be built up on the same principle. Initially the extension of such patterns would lead only to flat "grids," but Wrinch has concluded that these structures should be geometrically capable of folding into structures which approximate an arrangement which we might expect to find in the globular proteins. On the cyclol hypothesis, a limited number of types of space-enclosing molecules can be constructed, forming one or more series. In particular one series,  $C_n$ , takes the form of truncated tetrahedra and consists of  $72n^2$  residues; these numbers might be modified if imino-acid residues are present (reference 31c, p. 516). It must be remembered that all these structures are, as yet, hypothetical. No substances with a cyclol structure have yet been synthesized, and their chemical properties can only be inferred by analogy with other similar compounds.

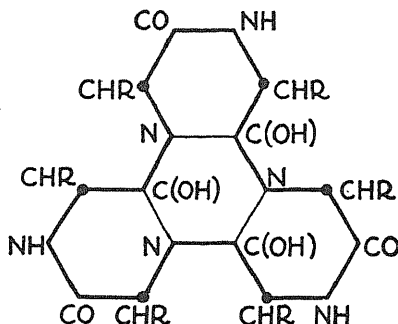
Wrinch has applied this reasoning in detail to the structure of pepsin<sup>32</sup> and insulin<sup>33</sup>. Both of these proteins, if the cyclol hypothesis were correct, should belong to the class  $C_2$  with 288 amino acid residues; but revealing,

<sup>31</sup> Wrinch, D. M., (a) *Nature*, 137, 411 (1936); (b) 138, 241 (1936); (c) 139, 972 (1937); (d) *Proc. Roy. Soc., London*, A160, 59 (1937); (e) 161, 505 (1937). Speculations of a less precise character have been advanced by Carl F. Kraft in two pamphlets: "Spiral Structure as the Basis of Life" and "Cyclol Structure as the Basis of Life" (Science Press, Philadelphia, Pa.).

of course, marked differences in crystal structure and chemical composition.

Later Wrinch and Langmuir<sup>34</sup> attempted to prove from the x-ray data of Crowfoot<sup>35</sup> on insulin that this molecule has the cyclol structure  $C_2$ . The proof has been questioned by several workers in the x-ray field,<sup>36, 37, 38</sup> who have concluded that the data available, and especially the methods of structure determination applicable to the data at present, are still quite inadequate to determine even the general pattern of the atomic arrangement in such complex systems. In the words of Fankuchen, "Given a set of x-ray intensities, we can assume a structure based on some theory and then test our theory by a comparison of observed and computed intensities. For proteins, however, this testing must be done very carefully with one eye on the character of the data and another on the theory to be tested. Thus, if the x-ray data are limited in scope (as is often the case

Figure 7.  
A "cyclol 6" molecule. From  
Wrinch, D. M., Proc. Roy. Soc.  
London, A161, 505 (1937).



for proteins) agreement between observed and computed intensities can say nothing about any fine structure of the theory and indeed need not even be a confirmation of the large-scale details of the structure being tested" (reference 39, p. 161).

The cyclol theory has also been criticized on chemical grounds, and on the ground that cyclol structures should be thermodynamically far more unstable than corresponding peptide chains—arguments set forth in detail by Pauling and Niemann<sup>40</sup>. These authors, and also Neurath<sup>17</sup> and Huggins<sup>41</sup>, have concluded that the side chains of the amino acid residues could not pack together as closely as demanded by the cyclol theory; and

<sup>34</sup> Wrinch, D. M., and Langmuir, I. J., *Proc. Roy. Soc. London*, **60**, 2247 (1938).

<sup>35</sup> Crowfoot, D. A., *Proc. Roy. Soc. London*, **161**, 505 (1937).

<sup>36</sup> Bragg, W. L., *Nature*, **143**, 100 (1939).

<sup>37</sup> Bragg, W. L., *Nature*, **143**, 100 (1939).

<sup>38</sup> Bragg, W. L., *Nature*, **143**, 100 (1939).

<sup>39</sup> Fankuchen, I., and Rilev, D., *ibid.*, **143**, 100 (1939).

<sup>40</sup> Pauling, L., and Niemann, O., *Proc. Nat. Acad. Sci.*, **14**, 100 (1918).

<sup>41</sup> Huggins, R. S., *Proc. Nat. Acad. Sci.*, **14**, 100 (1918).



Huggins has proposed, as an alternative, some ingenious models derived by folding of extended zigzag polypeptide chains, and held together by hydrogen bonds. Such models avoid the metrical difficulties associated with the cyclol theory, but the interpretation of the x-ray diffraction patterns of the globular proteins is still not sufficiently advanced to allow any decision concerning their validity though it is to be hoped that a succession of stimulating hypotheses comparable to the "cyclol theory" will continue to be advanced.

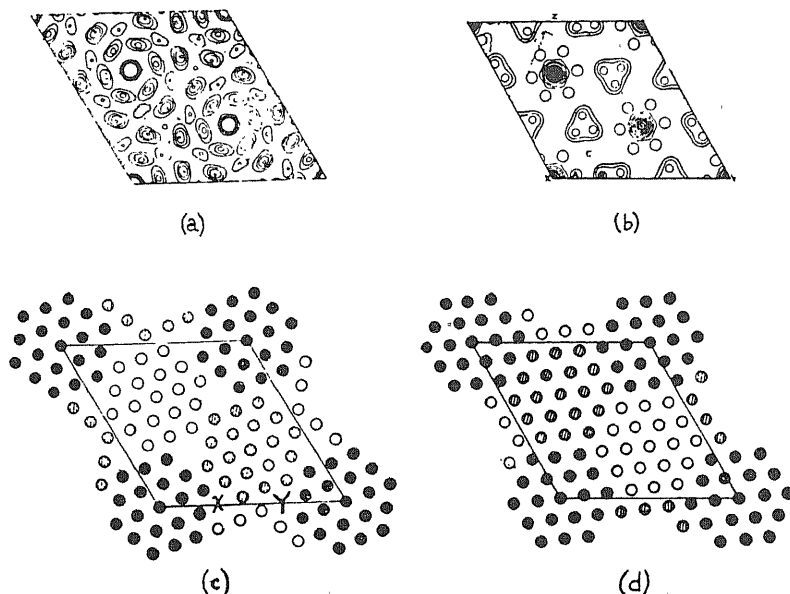


Figure 8. Patterson projections on the basal plane of the structure of (a) wet and (b) dry insulin. Arrangement of hexagonal arrays of points: (c) expanded (cf. wet insulin) and (d) closepacked (cf. dry insulin). From Crowfoot, D., *Chem. Rev.*, 28, 215 (1941).

#### Arrangement of Protein Molecules in Crystals: Patterson Projections

As yet it is impossible to deduce interatomic distances, within a protein molecule or between adjoining molecules, from x-ray data. Crowfoot<sup>24a</sup>, however, has obtained highly suggestive evidence for the presence of definite molecules in the insulin crystals, and for the reorientation of these molecules when the crystals are dried. Fig. 8 (a and b) shows the Patterson projections on the basal plane of the structure of wet and dry insulin<sup>42</sup>. Bernal<sup>43</sup> has pointed out that all the observed peaks fall on a hexagonal net-

work, the axes of which lie at an angle to the crystallographic axes. The lower diagrams, c and d, in Fig. 8, show the pattern of this hexagonal network in wet and dry insulin. In each projection there is a group of eighteen peaks around the origin (centered around the corners of the parallelograms in Fig. 8, c and d) which have the same position, relative to each other, in both projections. The whole group at the origin is turned, relative to neighboring groups, through an angle of about  $6^\circ$  in the wet as compared with the dry structure. This might be explained by the hypothesis that "the insulin molecule itself has a structure in which the eighteen points of the network around the origin are occupied by units which are arranged in a close packed array, not only within one molecule, but also with reference to the unit structure of neighboring molecules. The change from dry to wet insulin then appears to involve an angular shift of the molecules from these close packed positions. Further, the new peak positions in the wet (0001) projection are not far from a second hexagonal network, which might again bring the unit points into close contact" (reference 24a, p. 225). On this interpretation, the gaps at X, Y and other points (Fig. 8c) would represent spaces between the molecules, filled with liquid of crystallization. Similar, although less clearly defined, relations have been pointed out by Crowfoot for hemoglobin and lactoglobulin crystals.

These interpretations of the data must be regarded as tentative. They afford suggestions, however, of the possible scope of x-ray methods as applied to proteins. The advance in methods of structure determination during the last ten years has been so great that we may look forward to future developments in this field which will yield a far deeper understanding of proteins than we possess today.<sup>44</sup>

<sup>44</sup> For very recent developments of x-ray diffraction and protein structure, see Astbury, W. T., *J. Chem. Soc.*, p. 337 (1942); Hughes, E. W., *Review of Biochemistry*, 11, 27 (1942). A brief report on the crystal structure of glycylglycine ( $\beta$ -form) is given by Hughes, E. W., and Moore, W. J., *J. Am. Chem. Soc.*, 64, 2236 (1942).

## Chapter 15

# The Elementary and Amino Acid Composition of the Proteins

BY EDWIN J. COHN

Knowledge of the structure of proteins might well be expected to yield knowledge of the special functions which we know individual proteins subserve both in morphology and in physiology, as elements of structure or bearers of immunity, as hormones or enzymes, as the molecules to which certain of the vitamins appear to be attached as prosthetic groups, or as viruses. This diversity in function suggests an underlying diversity of structure, certainly in the finer details of structure.

Of the fine structure of the proteins little is yet known. We know that upon hydrolysis proteins yield amino acids; that the amino acids are at least in large part joined to each other in the vast protein molecule by peptide linkage; that many of the same amino acids have been isolated from widely different proteins; and that in many proteins, and even in certain enzymes and hormones<sup>1</sup>, there is no evidence that the proteins are constituted of any elements of structure other than amino acids. Furthermore, it is of great significance that the amino acids are all  $\alpha$ -amino acids (or  $\alpha$ -imino acids in two cases), and that all belong to the same configurational group, that is, that the four unlike radicals are invariably attached to the  $\alpha$ -carbon atom in the same order (see Chapter 13).

### Discovery of Amino Acids

The first of the amino acids were discovered a little over a century ago.\* Leucine and glycine were isolated from proteins in 1819 and 1820 by Proust<sup>2</sup> and by Braconnot<sup>3</sup>. Though cystine had previously been described by Wollaston in 1810<sup>4</sup>, it was not isolated from a protein until 80 years later<sup>5, 6</sup>. With the exception of tyrosine, which was found by Liebig in 1846<sup>7</sup>, no new amino acid was isolated from a protein for 45 years, or

<sup>1</sup> Cohn, E. J., *Bull. New York Acad. Med.*, **15**, 639 (1939) (Harvey Lecture).

\* For a history of the discovery of the amino acids, see the preface to the *Biochemical Handbook*, by E. J. Cohn, 1939, published by the National Academy of Sciences, Washington, D. C.

TABLE 1. THE ISOLATION OF AMINO ACIDS FROM PROTEIN HYDROLYSATES

Date	Amino acid	Isolated from protein by	
1819	Leucine	Proust	(1)
1820	Leucine	Braconnot	(2)
1820	Glycine	Braconnot	(2)
1846	Tyrosine	Liebig	(3)
1849	Tyrosine	Bopp	(4)
1865	Serine	Cramer	(5)
1866	Glutamic acid	Ritthausen	(6)
1868	Aspartic acid*	Ritthausen	(9)
1869	Aspartic acid*	Ritthausen (7) and Kreussler	(8)
1875	Alanine	Schützenberger and Bourgeois	(10)
1888	Alanine	Weyl	(11)
1881	Phenylalanine	Schulze and Barbieri	(12)
1889	Lysine	Drechsel	(13)
1890	Cystine†	Külz	(14)
1899	Cystine†	Mörner	(15)
1895	Arginine‡	Hedin	(16)
1896	Histidine	Kossel	(17)
1896	Histidine	Hedin	(18)
1901	Proline	Fischer	(19)
1901	Valine§	Fischer	(19)
1901	Tryptophane	Hopkins and Cole	(20, 20a)
1902	Oxyproline	Fischer	(21)
1903	Isoleucine¶	Ehrlich	(22, 23)
1918	Hydroxyglutamic acid	Dakin	(24)
1922	Methionine	Mueller	(25)
1936	Threonine	Meyer and Rose	(26)

- (1) Proust, M., *Ann. chim. et phys.*, **10**, 29 (1819).  
 (2) Braconnot, H., *Ann. chim. et phys.*, **13**, 113 (1820).  
 (3) Liebig, J., *Ann.*, **57**, 127 (1846).  
 (4) Bopp, F., *Ann.*, **69**, 16 (1849).  
 (5) Cramer, E., *J. prakt. Chem.*, **96**, 76 (1865).  
 (6) Ritthausen, H., *J. prakt. Chem.*, **99**, 6, 454 (1866).  
 (7) Ritthausen, H., *J. prakt. Chem.*, **106**, 445: 107, 218 (1869).  
 (8) Kreusler, W., *J. prakt. Chem.*, **107**, 240 (1869).  
 (9) Ritthausen, H., *J. prakt. Chem.*, **103**, 233 (1868).  
 (10) Schützenberger, P., and Bourgeois, A., *Compt. rend. Acad. Sci.*, **81**, 1191 (1875).  
 (11) Weyl, T., *Ber. chem. Ges.*, **21**, 1407 (1888).  
 (12) Schulze, E., and Barbieri, J., *Ber. chem. Ges.*, **14**, 1785 (1881).  
 (13) Drechsel, E., *J. prakt. Chem.*, **N. F.** **39**, 425 (1889).  
 (14) Külz, E., *Z. Biol.*, **27**, 415 (1890).  
 (15) Mörner, K. A. H., *Z. physiol. Chem.*, **28**, 595 (1899).  
 (16) Hedin, S. G., *Z. physiol. Chem.*, **20**, 186 (1895).  
 (17) Kossel, A., *Z. physiol. Chem.*, **22**, 176 (1896-97).  
 (18) Hedin, S. G., *Z. physiol. Chem.*, **22**, 191 (1896-97).  
 (19) Fischer, E., *Z. physiol. Chem.*, **33**, 151 (1901).  
 (20) Hopkins, F. G., and Cole, S. W., *J. Physiol.*, **27**, 418 (1901-02).  
 (20a) Hopkins, F. G., and Cole, S. W., *Proc. Roy. Soc. (London)*, **68**, 21 (1901).  
 (21) Fischer, E., *Ber. chem. Ges.*, **35**, 2660 (1902).  
 (22) Ehrlich, F., *Z. Ber. d. deutsch. Zucker Ind.*, **53**, 809 (1903).  
 (23) Ehrlich, F., *Ber. chem. Ges.*, **37**, 1809 (1904).  
 (24) Dakin, H. D., *Biochem. J.*, **12**, 290 (1918).  
 (25) Mueller, J. H., *Proc. Soc. Exptl. Biol. and Med.*, **19**, 161 (1922).  
 (26) Meyer, C. E., and Rose, W. C., *J. Biol. Chem.*, **115**, 721 (1936).

\* Previously discovered by Plisson in 1827 [Plisson, A., *Ann. chim. et phys.*, **36**, 175 (1827)].

† Previously discovered by Wollaston in 1810 [Wollaston, W. H., *Phil. Trans. Roy. Soc.*, **223** (1810)].

‡ Previously discovered by Schulze and Steiger in 1886 [Schulze, E., und Steiger, E., *Ber. chem. Ges.*, **19**, 1177 (1886)].

§ Previously discovered by v. Gorup-Besanez in 1856 [v. Gorup-Besanez, E., *Ann. Chem.*, **98**, 1 (1856)].

¶ The evidence is now very strong that norleucine (α-amino caproic acid) is present in spinal cord protein. This was first claimed by H. H. Koster, *Die chemische Konstitution des Gehirns des Menschen und der Tiere*, Fr. Pietzker, Tübingen, 1901) in 1901. He isolated a substance which he recognized as an isomer of leucine. His product was probably impure. Twelve years later Abderhalden and Weil [*Z.*

until 1865. In 1864 only three amino acids had thus been recognized among the decomposition products of proteins; in 1904 seventeen amino acids were known. In this intervening 40-year period, a new amino acid was discovered every few years and the main constituents of the protein molecule were recognized. Although more than 35 years have elapsed since 1904, a period nearly as long as that which saw the discovery of fifteen new amino acids, few further additions to our knowledge have been made.

**Sulfur-containing Amino Acids.** Two of the most important amino acids isolated from proteins in recent years have been discovered as a result of the increasing recognition of the importance of amino acids for nutrition. Studies of the amino acids necessary for growth and development, the pioneer work for which was largely carried out by Osborne and Mendel, have more recently yielded the new sulfur-containing amino acid, methionine, and the hydroxy-amino-butyric acid, threonine.

Methionine<sup>8</sup> was isolated from casein by Mueller in the course of a study of bacterial metabolism. Isolation was followed by synthesis and the establishment of its structure<sup>9a</sup> and also by fairly satisfactory methods for its estimation in proteins<sup>9, 10, 11</sup>.

Liebig and Mulder<sup>12</sup> knew that, besides nitrogen, carbon, hydrogen and oxygen, most proteins contain small amounts of sulfur and of phosphorus. The elementary composition of these proteins was so nearly alike that only the analysis for sulfur and phosphorus showed wide variations. The nature of the structures in which these elements occur in proteins is further considered below. (Table 2).

**Phosphorus and the Hydroxyamino Acids.** The nutritional investigations of Osborne and Mendel were for the most part carried out with proteins the amino acid composition of which they had estimated. The difficulties of obtaining adequate supplies of amino acids so that nutritional experiments could be carried out with purified amino acid mixtures have only recently been overcome. In the course of an investigation with this method of the amino acids necessary for growth, W. C. Rose has proved that  $\alpha$ -amino- $\beta$ -hydroxybutyric acid or threonine is definitely a protein constituent<sup>13</sup>. This is the more interesting because the claim<sup>14</sup> that  $\alpha$ -amino-*n*-butyric acid is a protein constituent is not yet substantiated and also because threonine is the second member of the series of which serine,  $\alpha$ -amino- $\beta$ -hydroxypropionic acid, is the first.

<sup>8</sup> Mueller, J. H., *Proc. Soc. Exp. Biol. and Med.*, **19**, 161 (1922).

<sup>9a</sup> Barger, G., and Coyne, F. P., *Biochem. J.*, **22**, 1417 (1928).

<sup>9</sup> Baernstein, H. D., *J. Biol. Chem.*, **97**, 663, 669 (1932); **106**, 451 (1934); **115**, 25, 33 (1938).

The phosphorus revealed by the elementary analyses of proteins has never been found as a part of an amino acid. Rather it would appear to be bound to hydroxyl groups such as that of serine through ester linkage. The investigations of Rimington and Kay<sup>15, 16</sup> appear to indicate that the phosphorus in these proteins can be quantitatively removed, and further investigations suggest that the phosphorus may be reintroduced into the molecules "by means of phosphorus oxychloride in a manner analogous to the Schotten-Baumann reaction" (reference 17, p. 280). "Phosphorus appears to enter the protein by means of an ester linkage between phos-

TABLE 2. ELEMENTAL COMPOSITION OF PROTEINS

Composition							Protein	Formula							Molecular weight
C	H	N	O	S	Fe	P		C	H	N	O	S	Fe	P	
According to Mulder in 1838* (1)															
54.56	6.90	15.72	22.13	0.36		0.33	Fibrin	400	620	100	120	1		155,692	
54.48	7.01	15.70	22.00	0.38		0.43	Egg albumin	400	620	100	120	1		155,692	
54.84	7.09	15.82	21.23	0.68		0.33	Serum albumin	400	620	100	120	2		155,893	
According to Osborne in 1902† (2)															
52.75	7.10	15.51	23.02	1.616			Egg albumin	696	1125	175	220	8		15,703	
55.23	7.26	16.13	20.78	0.600			Zein	736	1161	184	208	3		15,993	
52.72	6.86	17.66	21.73	1.027			Gliadin	685	1068	196	211	5		15,568	
52.99	7.01	15.93	22.14	1.930			Serum albumin†	662	1051	171	207	9		14,989	
54.64	7.09	17.38	20.16	0.39	0.335		Hemoglobin†	752	1181	207	210	2	1	16,655	
54.57	7.11	16.38	21.03	0.568	0.336		Hemoglobin‡	752	1181	207	210	3	1	16,667	
52.68	6.83	16.91	22.48	1.10			Albumin§	685	1068	196	211	5		14,708	
51.50	7.02	18.69	21.91	0.88			Edestin	624	1021	193	199	4		14,523	
53.13	7.06	15.78	22.37	0.80		0.86	Casein	708	1130	180	224	4	4	15,982	

\* For a critical study see Laskowsky (Ann. 58, 129 (1846)). Mulder used atomic weights based on O = 100.  
 † O. . . . . the earlier literature in this paper. See, also, Osborne: The Vegetable Proteins, 2nd ed., p. 10, Green and Co. (1924).

‡ Of the horse.

§ Of the dog.

(1) Mulder, G. J., *Ann. Chem.*, **28**, 73 (1838).

(2) Osborne, T. B., *J. Am. Chem. Soc.*, **24**, 140 (1902).

phoric acid and some hydroxyl group" (reference 17, p. 281), and, in the case of casein, isolation of diverse phosphorus containing polypeptides and peptones have been reported<sup>18</sup>, together with studies of their constitution<sup>19, 20, 21, 21a, 21b</sup>.

Whereas serine and threonine have now been repeatedly demonstrated

<sup>15</sup> Rimington, C., and Kay, H. D., *Biochem. J.*, **20**, 777 (1926).

<sup>16</sup> Rimington, C., *Biochem. J.*, **21**, 204 (1927).

<sup>17</sup> Rimington, C., *Biochem. J.*, **21**, 272 (1927).

<sup>18</sup> Rimington, C., *Biochem. J.*, **21**, 1179 (1927).

<sup>19</sup> . . . . .

to be constituents of proteins, other basic and hydroxyamino acids have repeatedly been reported, although the observations have seldom been verified. Thus in 1907 Abderhalden and Kempe<sup>22</sup> announced a new amino acid, oxytryptophane. Later, however, the preparations were found to have been mixtures of well-known amino acids<sup>23, 23a</sup>. There are also claims for the presence of hydroxyvaline<sup>24</sup> and hydroxyaspartic acid<sup>25</sup> in protein hydrolysates.<sup>26</sup>

In 1918 Dakin<sup>27</sup> isolated an amino acid from casein which he considered to be  $\beta$ -hydroxyglutamic acid. Although the existence of this amino acid in proteins has been doubted by some investigators, its presence has been reported by others and consideration has recently been given<sup>14</sup> to certain of the difficulties involved in its isolation, and to the controversy regarding the evidence that this amino acid is a normal constituent of proteins. Recently Dakin<sup>27a</sup> has pointed out that the properties of his original preparation do not correspond to those to be expected of a  $\beta$ -hydroxy amino acid. The possibility that it is another isomer remains to be determined.

Very recently hydroxylysine has been reported as a constituent of gelatin by Van Slyke<sup>28</sup> and its structure suggested to be either  $\delta$ -hydroxylysine or  $\alpha$ ,  $\delta$ , diamino,  $\epsilon$ -hydroxy-caproic acid<sup>28a</sup>. A method of estimating the hydroxyl group by oxidation with periodate has also been adopted for its estimation and further work with this procedure may well render it possible either to substantiate or reject the many claims for hydroxy amino acids that have been made. Recent work of Nicolet<sup>28b</sup> and of Martin and Synge<sup>28c</sup> is most promising in this connection.

**Rare Amino Acids.** Besides those amino acids which are constituents of most proteins, there are certain ones which have been isolated from but few proteins. Their isolation has none the less been confirmed and their importance established.

Thus the iodine, long associated with the thyroid gland and with thyroglobulin, is a constituent of diiodotyrosine and also of thyroxine. Although Bubnow in 1884<sup>29</sup> identified the thyroid colloid as protein in nature, it was not until 1896 that Baumann<sup>30</sup> and Hutchinson<sup>31</sup> studied thyroglobulin as an iodoprotein. Meanwhile, the isolation of iodogorgoic acid by Drechsel

<sup>22</sup> Abderhalden, E., and Kempe, M., *Z. Physiol. Chem.*, **52**, 207 (1907).

<sup>23</sup> Abderhalden, E., and Sichel, H., *Z. Physiol. Chem.*, **138**, 108 (1924).

<sup>23a</sup> Abderhalden, E., and Sichel, H., *Z. Physiol. Chem.*, **144**, 80 (1925).

<sup>24</sup> Sichel, H., and Abderhalden, E., *Proc. Roy. Soc. London*, **B98**, 58 (1925).

<sup>25</sup> Sichel, H., and Abderhalden, E., *Proc. Roy. Soc. London*, **B98**, 58 (1925).

<sup>26</sup> Sichel, H., and Abderhalden, E., *Proc. Roy. Soc. London*, **B98**, 58 (1925).

<sup>27</sup> Dakin, H. D., *J. Biol. Chem.*, **140**, 847 (1941).

<sup>27a</sup> Dakin, H. D., Hiller, A., Dillon, R. T., and MacFadyen, D., *Proc. Soc. Exp. Biol. and Med.*, **38**, 548 (1938).

<sup>28</sup> Van Slyke, D. D., Hiller, A., MacFadyen, D. A., Hastings, A. B., and Klemperer, F. W., *J. Biol. Chem.*, **133**, 287 (1940).

<sup>28a</sup> Van Slyke, D. D., Hiller, A., MacFadyen, D. A., Hastings, A. B., and Klemperer, F. W., *J. Biol. Chem.*, **133**, 287 (1940).

<sup>28b</sup> Nicolet, B. H., and Shinn, T. A., *ibid.*

<sup>28c</sup> Martin, R. M., and Synge, A. L. C., *ibid.*

in 1895<sup>32</sup> had suggested the existence of iodine-containing amino acids, and the subsequent identification of this substance as 3,5-diiodotyrosine in 1905-9 by Wheeler and Jamieson<sup>33</sup> and Wheeler and Mendel<sup>34</sup> led to a systematic search for the iodo-amino acid, not only in thyroglobulin, but also in various iodine-containing materials, both natural and synthetic. Thus, in a series of papers about 1911, Oswald<sup>35</sup> described the isolation of the compound from several synthetic iodoproteins, *e.g.*, iodinated egg albumin, and iodinated casein as well as from natural gorgonin and spongin. These results intensified the suspicion that diiodotyrosine is a natural constituent of thyroglobulin. Satisfactory proof of this was ultimately secured by Harington and his collaborators<sup>36, 37, 38</sup> who prepared the substance from enzymic hydrolysates of thyroglobulin.

In the meantime, after two decades of disappointing research, Kendall in 1915<sup>39</sup> had reported the isolation of crystalline racemic thyroxine, the bearer of the characteristic endocrine activity of the thyroid. Its constitution was finally established by Harington and Barger in 1927<sup>40</sup>, and the natural *l*-form was isolated from enzymic hydrolysates of thyroglobulin by Harington and Salter in 1930<sup>38</sup>.

In addition, Ludwig and Mutzenbecher have isolated<sup>41</sup> racemic monoiodotyrosine from alkaline hydrolysates of iodinated casein.

In the course of their synthesis of thyroxine, Harington and Barger<sup>40</sup> prepared 3,5-diiodothyronine, and subsequently Harington and his collaborators prepared various homologs of thyroxine containing other halogens. (See also Niemann, *et al.*<sup>42</sup>)

Although iodinated histidine has not yet been itself prepared, there are good grounds for suspecting the existence of both mono-iodohistidine and diiodohistidine in iodinated proteins. Indeed, Pauly has prepared tetraiodohistidine<sup>43</sup> anhydride and paranitrobenzoyl diiodohistidine, which serve as structural analogs. Furthermore, Bauer and Strauss<sup>44</sup> have made a study of the step-wise iodination of globin, which indicates that this process can be described adequately in the following terms: (a) initial iodination of tyrosine followed by (b) iodination of histidine (carbon linkage after intermediate nitrogen combination), followed by (c) iodination of the regenerated NH group of histidine to form the diiodo derivative.

Closely related chemically to diiodotyrosine is dibromotyrosine which

<sup>32</sup> Drechsel, E., *Zentralbl. Physiol.*, **9**, 705 (1895).

<sup>33</sup> Wheeler, H. L., and Jamieson, G. S., *Am. Chem. J.*, **33**, 365 (1905).

<sup>34</sup> Wheeler, H. L., and Mendel, L. B., *J. Biol. Chem.*, **7**, 1 (1909).

<sup>35</sup> Oswald, A. Z., *Pharmaz. Chem.*, **71**, 200 (1910-11); **74**, 290 (1911); **75**, 353 (1911).

<sup>36</sup> Harington, J. D., and Barger, J. F., *Biochem. J.*, **23**, 373 (1929).

<sup>37</sup> Harington, J. D., and Barger, J. F., *Biochem. J.*, **25**, 1032 (1931).

<sup>38</sup> Harington, J. D., and Salter, W. T., *Biochem. J.*, **24**, 456 (1930).



was isolated by Mörner<sup>45</sup> from the skeleton of *Primnoa lepadifera*. Its wider distribution remains to be demonstrated.

In addition to these rare amino acids of animal origin, Vickery<sup>14</sup> lists the following "amino acids known as plant constituents that may possibly be expected to be found in proteins": Thiolhistidine found in ergot as its betaine, ergothioneine<sup>46</sup>; dihydroxyphenylalanine found in bean seedlings and probably widely distributed<sup>47, 48</sup>; citrulline found in watermelon tissue and probably of metabolic significance in urea formation in animals<sup>49</sup>; canavanine found in certain beans<sup>50</sup>; and djenkolic acid found in djenkol bean<sup>51</sup>.

It is very probable that other amino acids of special structure and significance will be found in the investigations of diverse kinds that are in progress on the rarer and less well known proteins. Some of these proteins, though present in small amounts, may perform quite as important physiological functions as those present in large amounts. The latter, however, have thus far been more readily available in adequate amounts for studies, especially for hydrolytic studies, of their amino acid composition.

### Classes of Amino Acids

The amino acids may, for convenience, be divided into classes each of which contains certain characteristic groups and confers specific properties on the protein molecule.

(1) The first class may be considered the monoamino-monocarboxylic  $\alpha$ -amino acids whose side chains differ only in the number of  $\text{CH}_2$  groups that they contain and in the branching of the chain. If glycine, with no side chain, be considered the first member of this series the others are alanine, valine, leucine and isoleucine, and if they exist in proteins, norvaline and norleucine.

(2) The second group contribute pyrrolidine rings rather than paraffin side chains and are imido rather than  $\alpha$ -amino acids. Proline and hydroxyproline have been isolated from many proteins, often in large amounts. In hydroxyproline the nonpolar groups are situated between the hydroxyl groups and the peptide chain.

(3) The third group consists of those amino acids whose side chains contain benzene rings; phenylalanine, tyrosine, diiodotyrosine and thyroxine. In tyrosine, the benzene ring is situated between polar groups; and the phenolic hydroxyl group, as we have seen, has pronounced acid properties. Tyrosine gives rise also to many of the chromogenic properties which have often been used as tests for the presence of protein and this substance has also been shown to play an important role in immunological reactions.

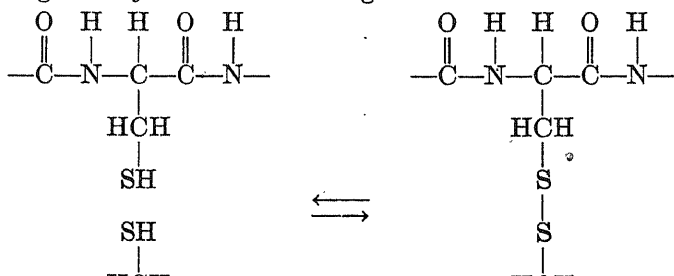
(4) Closely related to tyrosine, in that its presence in proteins is generally determined by virtue of its color reactions, is tryptophane. Tryptophane has no amphoteric properties when bound in peptide linkage, but contains the indole ring as a characteristic configuration.

(5) The amino acids serine and threonine contain hydroxyl groups whose acidity is so feeble as not to contribute to the amphoteric properties of the protein. This is true also for hydroxyproline, hydroxylysine and hydroxyglutamic acid, but these three amino acids may also be considered in connection with their other characteristic groups, namely, the pyrrolidine ring, the  $\epsilon$ -amino group and the  $\gamma$ -carboxyl group of these amino acids.

(6) Only three amino acids containing two carboxyl groups have yet been recognized as protein constituents; they are aspartic, glutamic and hydroxyglutamic acid. The second carboxyl groups of these acids are present in the native protein, in part as free acid, in part as amide groups. Asparagine and glutamine have been isolated from proteins<sup>14</sup>, and it may be assumed that most of the ammonia produced by acid hydrolysis represents the nitrogen of these amide groups. Threonine and serine, however, when in peptide linkage, are very labile in alkali solution and split off considerable quantities of ammonia which might thus be erroneously considered as "amide nitrogen"<sup>51a</sup>.

(7) In contrast to the three dicarboxylic acids upon which the acid properties of proteins depend are the bases, histidine, arginine and lysine, which confer basic properties. The side chain groups of these dibasic amino acids are all different, being respectively the imidazol, the guanidino and the  $\epsilon$ -amino group.

(8) Finally there are the sulfur-containing amino acids, methionine, cystine and cysteine. The relative amounts of cystine and cysteine in proteins have been repeatedly studied in recent years, but need not be considered here. In so far as cystine is present, the possibility exists that separate peptide chains, following oxidation of adjacent sulfhydryl groups, are held together by sulfur-sulfur linkage.



Inasmuch as cysteine is present the sulfhydryl groups should contribute acid properties to the molecule, comparable in strength to those derived from phenolic hydroxyl groups. The possibility remains of an equilibrium between S—S and S—H groups in the protein involving oxidation and reduction and is considered elsewhere in this book.

The sulfur of methionine is situated between a CH<sub>2</sub> and a CH<sub>3</sub> group, is not alkali labile, and is thus readily distinguished from cystine or cysteine sulfur.

### Analysis of Amino Acids

The accuracy with which analysis may be made for amino acids of these various classes differs widely. The methods of analysis are not considered here in detail, but only the results of the analyses upon which we must depend for our notion of the free groups of proteins. Certain analyses can be made, not only for the individual amino acid, but also for the class of which it is a member, and in these cases greater confidence may be felt in the accuracy of the results. Thus, if the sum of the methionine and cystine analyses is in agreement with what might be expected from the behavior of the sulfur, we may have increased confidence in both procedures. (Table 3).

**Sulfur-containing Amino Acids.** The sulfur of proteins is of two kinds as was clearly perceived by the analytical chemists of the late nineteenth century, such as Schultz<sup>52</sup>, Krüger<sup>53</sup> and T. B. Osborne<sup>54, 55</sup>. Krüger was the first to present accurate analytical data for the total and the alkali-labile sulfur. Moreover, he pointed out that, for egg albumin and for fibrin, the ratio of total sulfur to labile sulfur was, respectively, 4:1 and 3:1\*.<sup>53</sup> The sulfide or labile sulfur therefore represents an integral fraction of the total sulfur. Osborne<sup>54, 55</sup> in 1902 estimated, on the basis of the results in Table 2, that the minimal molecular weights of many proteins were in the neighborhood of 15,000, estimates borne out by recent physicochemical studies.

At the beginning of this century, there were no satisfactory analytical methods for determining cystine, and the other sulfur-containing amino acid thus far isolated from a protein hydrolysate, methionine, had not been discovered<sup>56, 58</sup>. The sulfur in cystine is relatively labile and is liberated as

<sup>52</sup> Schulz, F. N., *Z. Physiol. Chem.*, **25**, 16 (1898).

<sup>53</sup> Krüger, A., *Arch. ges. Physiol. (Pflügers)* **43**, 244 (1888).

<sup>54</sup> Osborne, T. B., *J. Am. Chem. Soc.*, **24**, 140 (1902).

<sup>55</sup> Osborne, T. B., *Proteins*, Longmans, Green and Co., London and New York (1909).  
 str.: and repeating ... for the occurrence of amino acids in proteins has since been demon-  
<sup>56</sup> Bergmann, M., *Chem. Rev.*, **22**, 423 (1938).

sulfide ion in hot alkaline solution. This sulfur is even more labile when the amino acid is held in peptide linkage in the protein molecule<sup>59, 60, 61</sup>

The sulfide sulfur of the protein has been supposed to represent, at least in part, the sulfur either in the amino acid cystine or cysteine. Cysteine contains one atom of sulfur, while cystine contains two. Whatever the nature of the sulfide sulfur, it represents an integral part of the protein molecule, and the quantities that analyses reveal can be used in estimating the minimal molecular weights of proteins in the same manner as can their iron, copper, or total sulfur content<sup>62</sup>.

The total sulfur content of most proteins is greater than their sulfide sulfur content. This would indicate the presence of yet another sulfur-containing constituent of the protein. The character of this non-labile sulfur was demonstrated by Mörner who obtained methyl sulfonic acid on treating proteins with nitric acid. This author<sup>63</sup> further pointed out that the amount of methyl sulfonic acid appeared to be inversely proportional to the cystine content in the cases of casein and wool. It was not, however, until 1923 that a second sulfur-containing amino acid was isolated from a protein hydrolysate. In that year Mueller<sup>8, 58</sup> isolated methionine from casein and it is doubtless significant that in casein there is not more than 0.1% of sulfide sulfur, but approximately 0.8% of total sulfur and over 3% of methionine.

Both cystine and methionine can be estimated with some accuracy in proteins. There are a series of analytical methods for cystine due to Folin and his co-workers<sup>64</sup>, to Sullivan<sup>65</sup>, to Okuda<sup>66</sup> to Vickery and White<sup>67</sup>, and to others.

The estimation of methionine depends on the reaction of this amino acid with hot hydriodic acid whereby the S—CH<sub>3</sub> bond is ruptured and homocysteine is formed. Baernstein<sup>9</sup> has developed two quite accurate methods based on this reaction for determination of methionine in proteins.

In many proteins, an exception being crystalline insulin which has been carefully studied by du Vigneaud and Miller<sup>68</sup>, the combined cystine and methionine sulfur at first appeared not to equal the total sulfur. In the case of insulin, the introduction of Clarke's formic-hydrochloric acid technique of hydrolysis, used by du Vigneaud and Miller and by Sullivan<sup>69</sup>,

<sup>59</sup> Brand, E., and Sandberg, M., *J. Biol. Chem.*, 70, 381 (1926).

<sup>60</sup> Bergmann, M., and Stather, F., *Ann.*, 448, 32 (1926).

<sup>61</sup> Mörner, K. A. H., *Z. Physiol. Chem.*, 34, 207 (1902).

<sup>62</sup> Cohn, E. J., Hendry, L. J., and Prentiss, A. M., *J. Biol. Chem.*, 63, 721 (1925).

<sup>63</sup> Mörner, K. A. H., *Ann.*, 1914.

<sup>64</sup> Folin, D., *J. Biol. Chem.*, 53, 103 (1929).

<sup>65</sup> Sullivan, J. A., *J. Biol. Chem.*, 41, 1030 (1926).

<sup>66</sup> Okuda, Y., *J. Biochem. Japan*, 5, 201, 217 (1925).

<sup>67</sup> Vickery, H. R., and White, A., *J. Biol. Chem.*, 60, 701 (1923).

TABLE 3. SULFUR CONTAINING AMINO ACIDS OF PROTEINS\*

	Egg albumin (1)	Insulin (1)	Zein	Gliadin	Casein	Edestin (1)	Gelatin	Hemo-globin (horse) (1)	Limulus hemo-cyanin	Myosin (rabbit)	Fibrin	Serum albumin	Serum albumin (horse)	Chymo-trypsi-nogen	Thyro-globulin
.....	1.616 (2)	3.34 (4, 5)	0.60 (2)	1.03 (2)	0.78 (3)	0.88 (2)	0.14 (9)	0.39 (11)	1.22 (13)	1.10 (7)	0.97 (8b)	1.89 (12)	1.68 (32)	1.48 (33)	1.45 (34)
.....	0.491 (2)	0	0.21 (2)	0.62 (2)	0.09 (9)	0.346 (2)	0.03 (9)	0.19 (12)			0.38 (14)	1.28 (12)			
.....	1.125	0	0.39	0.41	0.69	0.534	0.11	0.20			0.59	0.61			
.....	1.78 (3)	12.5 (4)	0.91 (6)	2.40 (7)	0.42 (10)	1.36 (7)	0.17 (19)	1.02 (10)	1.94 (13)	0.77 (7)	1.51 (6)	5.71 (15)	6.28 (32)	4.59 (33)	4.30 (34)
.....	5.23 (3)	0	2.35 (8a, c)	2.03 (8a)	3.25 (8)	2.39 (8b)		0.94 (10)	2.76 (13)	3.4 (7)	2.59 (8b)		0.00 (32)	1.22 (33)	1.31 (34)
Atoms or residues $\times 10^6$ per gram of protein															
.....	15.3	104	6.6	19.3	2.8	10.8	0.94	6.0			11.9	39.9			
.....	14.7	104	7.6	19.9	3.5	11.2	1.4	8.4	16.2	6.41	12.6	47.5	52.3	38.2	35.8
.....	35.1	0	12.1	12.8	21.5	16.7	3.43	6.2			18.4	19.0	0.0	8.2	8.8
.....	35.1	0	15.8	13.6	21.8	16.0		6.3	18.5	22.8	17.4				
.....	50.4	104	18.7	32.1	24.3	27.5	4.4	12.2	38.1	34.3	30.3	59.0	52.3	46.2	45.2
.....	49.7	104	23.4	33.5	25.3	27.3		14.7	34.7	29.2	30.0	60.2	52.3	46.4	44.6

..... in the upper part of this table are per cent.  
 ..... as cysteine whether present in this form or as cystine.  
 ..... is recrystallized; nearly carbohydrate free; N content 15.90%.  
 ..... in Table 3 should be compared with those given by Osborne (2) (see Table 2 of this chapter), by Cohn (16, 24, 29),  
 ..... and White (6), by Toennies (31) and by Baernstein (8).  
 ..... The discrepancy between the high total sulfur and the small estimates for cystine, and for lead blackening sulfur, is  
 ..... accounted for by the high methionine content. On the basis of the earlier determinations of sulfur and sulfide sulfur of Os-  
 ..... previously estimated 17 sulfur and 5 sulfide sulfur atoms in egg albumin, yielding molecular weights of 33,700 and 32,700  
 ..... The difference, 12 atoms of sulfur, is equal to the sulfur in the methionine.  
 ..... there are new estimates both of total and of alkali-labile sulfur. Baernstein (8) estimates 1.60% and Blumenthal and  
 ..... % total sulfur, in excellent agreement with Osborne's early value of 1.616%. Blumenthal and Clarke (9) give 0.43% as the  
 ..... sulfur. Five and six atoms of sulfide sulfur would on this basis yield molecular weights respectively of 37,300 and 44,700.  
 ..... and Quackenbush (10) give a value for total sulfur of 1.59%, which is in good agreement with Osborne's, and this and their  
 ..... er value for methionine, 4.85%, lead to a molecular weight of 36,000 to 37,000 on the basis of twelve residues per mole. The  
 ..... e, 1.78%, is moreover far greater than that previously reported yielding six cysteine residues and removing the discrepancy  
 ..... between sulfide sulfur and cystine.  
 ..... ll of the sulfur in insulin is presumably present as cystine, for the minimum molecular weight per sulfur atom is 958 and for  
 ..... e residues 961. On the basis of 36 sulfur atoms the molecular weights calculated from the sulfur and cystine determinations  
 ..... 34,488 and 34,596 respectively (10).

zein only a part of the total sulfur is present as sulfide sulfur. Blumenthal and Clarke (9) estimate 0.22% alkali labile sulfur, agreement with Osborne's earlier result of 0.212%. These values lead to minimum molecular weights of 14,570 and 15,270 and are consistent therefore with either 2, or probably 3, sulfide sulfur atoms in the zein molecule. The estimates of total considerably. Osborne's earlier value of 0.60% is in fair agreement with the newer estimate of total sulfur of 0.52% of Blumenthal and Clarke (9). These authors suggest however that 0.6% represented "pre-existing sulfate sulfur." The difference 0.36% is greater than their estimate of sulfide sulfur and far smaller than the estimates of Baernstein (8) which range from 0.93 to 0.97% as much as 11 or 12 sulfur atoms in the zein molecule. More probably the high percentage, 24.5, of total sulphur recovered by Baernstein (8d), represented a high proportion of sulfate sulfur impurities not present in the preparation of Osborne or Blumenthal and Clarke.

Osborne, Baernstein reports values varying from 2.25 to 2.58%, depending upon the source of the preparation and the method used. Osborne's value for sulfide sulfur has not been confirmed, but is in better agreement with the high value for cystine that Bailey (2, 40%) than with the values of Vickery and White (6), 2.06 and 2.15%. Osborne's value for sulfur is intermediate between Baernstein, 0.99%, and that of Bailey, 1.19%. The value for non-sulfide sulfur derived from it is also intermediate between the two, the methionine value of Bailey, 1.60%, and of Baernstein, 2.03%, though somewhat closer to the higher value of the latter than to the former.

Osborne's value, 0.619% (2), is adequate to account for the cystine in gliadin. Assuming 4 atoms of sulfide sulfur yields a minimum weight of 20,700 (16). The cystine percentage calculated on the assumption of 2 residues (4 of cysteine) in a molecule of this size is 0.619% (16). The value of Folin and Looney (17) was 2.32%, of Folin and Marenzi (18) 2.18%, and of Looney (19) 2.40%. The highest of White's values was 2.15%, whereas those of Sullivan and Hess (20) vary from 2.16 to 2.14%.

Hammarsten (21) estimated that casein contained 0.76% sulfur, Osborne (2) 0.80%. A large number of sulfur determinations made by various methods have since then been reported which range from the value 0.63% of Blumenthal and Clarke (9) to 0.90% of Plimmer and Lowndes. Osborne's earlier value of 0.80%, and other values of Baernstein of 0.80 and 0.83% are intermediate, as are the values reported by Blumenthal and Clarke and by Plimmer and Lowndes are appreciably lower than most of the other values.

Osborne's value, 0.619% (2), is adequate to account for the cystine in gliadin. Assuming 4 atoms of sulfide sulfur yields a minimum weight of 20,700 (16). The cystine percentage calculated on the assumption of 2 residues (4 of cysteine) in a molecule of this size is 0.619% (16). The value of Folin and Looney (17) was 2.32%, of Folin and Marenzi (18) 2.18%, and of Looney (19) 2.40%. The highest of White's values was 2.15%, whereas those of Sullivan and Hess (20) vary from 2.16 to 2.14%.

Hammarsten (21) estimated that casein contained 0.76% sulfur, Osborne (2) 0.80%. A large number of sulfur determinations made by various methods have since then been reported which range from the value 0.63% of Blumenthal and Clarke (9) to 0.90% of Plimmer and Lowndes. Osborne's earlier value of 0.80%, and other values of Baernstein of 0.80 and 0.83% are intermediate, as are the values reported by Blumenthal and Clarke and by Plimmer and Lowndes are appreciably lower than most of the other values.

Osborne's value, 0.619% (2), is adequate to account for the cystine in gliadin. Assuming 4 atoms of sulfide sulfur yields a minimum weight of 20,700 (16). The cystine percentage calculated on the assumption of 2 residues (4 of cysteine) in a molecule of this size is 0.619% (16). The value of Folin and Looney (17) was 2.32%, of Folin and Marenzi (18) 2.18%, and of Looney (19) 2.40%. The highest of White's values was 2.15%, whereas those of Sullivan and Hess (20) vary from 2.16 to 2.14%.

Hammarsten (21) estimated that casein contained 0.76% sulfur, Osborne (2) 0.80%. A large number of sulfur determinations made by various methods have since then been reported which range from the value 0.63% of Blumenthal and Clarke (9) to 0.90% of Plimmer and Lowndes. Osborne's earlier value of 0.80%, and other values of Baernstein of 0.80 and 0.83% are intermediate, as are the values reported by Blumenthal and Clarke and by Plimmer and Lowndes are appreciably lower than most of the other values.

TABLE 3—Concluded

bin. The values for sulfur and sulfide sulfur are those quoted by Osborne, whereas the values tentatively accepted for cystine are those of Kuhn, Birköfer and Quackenbush (10), which are far higher, indeed more than double, the earlier value for ckerly and White (6), 0.41, which yields the results of  $3.4 \times 10^{-5}$  residues of cystine per gram of hemoglobin as contrasted with imated from the report of Kuhn, Birköfer and Quackenbush. The agreement between the non-sulfide sulfur and the methio-oser.

Weber (27) reported a value for sulfur in myosin of the rabbit of 1%, and Todrick and Walker (28) of 0.94%. The minimum eight for rabbit myosin estimated from the sulfur content is 2.910, on the basis of Bailey's estimate, and Weber's lower yield eight of 3.410.

The value of Folin and Looney (17) for cystine, is three times as large as that now reported by Vickery and White (6), 1.51%, n (8), 1.58%, by Teruchi and Okabe (30), and by Bailey (7), for this protein. The average of the results reported by the former, l the values reported by these other investigators all fall between 1.48 and 1.64%. These results are consistent with the amount, postulated on the basis of the earlier sulfide sulfur determination of Krüger (14). If Krüger's value for total sulfur, 1.2%, is e molecules of non-sulfide sulfur become  $25.6 \times 10^{-5}$  per gram higher than Bailey's estimate of the methionine in fibrin. Since a particularly well defined protein, the disagreement is in the results from different laboratories and may not be significant. 1, however, has reported a lower value of total sulfur, 1.1%, and recently Baernstein (8) has reported a far lower value, 0.97%, s of sulfur per gram of fibrin. This result is in excellent agreement with the expectation from the recent estimates of cysteine ively accepted.

bumin. Preparations of serum albumin of the purity achieved by Hewitt (25) and by McMeekin (26) have not been analyzed es for sulfide sulfur and total sulfur of Schultz (12) are reported, though Blumenthal and Clarke (9) have since suggested far of 0.62 and 1.03% respectively. Whereas no methionine value is reported, the values for cystine, in which serum albumin is rich, are reported as 6.2% by Folin and Marenzi (18) and 5.71% by Sullivan (15). These results were on the same crystals prepared in this laboratory. Since Sullivan made a very careful comparative investigation of this preparation with different this result is tentatively accepted, though it, as well as other analytical evidence, should be repeated with the crystallbumin n purified by crystallization as sulfate (26).

B., *Annals New York Acad. Sci.*, 41, 87 (1941) Values assumed by ave generally been accepted.

- (18) Folin, O., and Marenzi, A. D., *J. Biol. Chem.*, 83, 103 (1929).
- (19) Looney, J. M., *J. Biol. Chem.*, 69, 519 (1926).
- (20) Sullivan, M. X., and Hess, W. C., *Publ. Health Rep., U.S.P.H.S.*, Suppl. 86 (1930).
- (21) Hammarsten, O., *Z. Physiol. Chem.*, 7, 227 (1882-83).
- (22) Plimmer, R. H. A., and Lowndes, J., *Biochem. J.*, 31, 51 (1937).
- (23) Kassel, B., and Brand, E., *J. Biol. Chem.*, 125, 435 (1938).
- (24) Cohn, E. J., *Ergbn. d. Physiol.*, 33, 781 (1933).
- (25) Hewitt, L. F., *Biochem. J.*, 30, 2229 (1936); *ibid.*, 32, 26 (1938).
- (26) McMeekin, T. L., *J. Am. Chem. Soc.*, 61, 2884 (1939).
- (27) Weber, H. H., *Ergbn. d. Physiol.*, 36, 109 (1934).
- (28) Todrick, A., and Walker, E., *Biochem. J.*, 31, 292 (1937).
- (29) Cohn, E. J., *Chem. Rev.*, 24, 203 (1939).
- (30) Teruchi, Y., and Okabe, L., *J. Biochem. Japan*, 8, 459 (1928).
- (31) Toennies, G., *Growth*, 1, 337 (1937).
- (32) Brand, E., and Kassel, B., *J. Biol. Chem.*, 141, 999 (1941).
- (33) Brand, E., and Kassel, B., *J. Gen. Physiol.*, 25, 157 (1941).
- (34) Brand, E., Kassel, B., and Heidelberger, M., *J. Biol. Chem.*, 128, xi (1939).
- (35) Meyer, K., *Cold Spring Harbor Symp. Quant. Biol.*, 6, 91 (1938).

- (18) Folin, O., and Marenzi, A. D., *J. Biol. Chem.*, 83, 103 (1929).
- (19) Looney, J. M., *J. Biol. Chem.*, 69, 519 (1926).
- (20) Sullivan, M. X., and Hess, W. C., *Publ. Health Rep., U.S.P.H.S.*, Suppl. 86 (1930).
- (21) Hammarsten, O., *Z. Physiol. Chem.*, 7, 227 (1882-83).
- (22) Plimmer, R. H. A., and Lowndes, J., *Biochem. J.*, 31, 51 (1937).
- (23) Kassel, B., and Brand, E., *J. Biol. Chem.*, 125, 435 (1938).
- (24) Cohn, E. J., *Ergbn. d. Physiol.*, 33, 781 (1933).
- (25) Hewitt, L. F., *Biochem. J.*, 30, 2229 (1936); *ibid.*, 32, 26 (1938).
- (26) McMeekin, T. L., *J. Am. Chem. Soc.*, 61, 2884 (1939).
- (27) Weber, H. H., *Ergbn. d. Physiol.*, 36, 109 (1934).
- (28) Todrick, A., and Walker, E., *Biochem. J.*, 31, 292 (1937).
- (29) Cohn, E. J., *Chem. Rev.*, 24, 203 (1939).
- (30) Teruchi, Y., and Okabe, L., *J. Biochem. Japan*, 8, 459 (1928).
- (31) Toennies, G., *Growth*, 1, 337 (1937).
- (32) Brand, E., and Kassel, B., *J. Biol. Chem.*, 141, 999 (1941).
- (33) Brand, E., and Kassel, B., *J. Gen. Physiol.*, 25, 157 (1941).
- (34) Brand, E., Kassel, B., and Heidelberger, M., *J. Biol. Chem.*, 128, xi (1939).
- (35) Meyer, K., *Cold Spring Harbor Symp. Quant. Biol.*, 6, 91 (1938).

enabled these workers to revise the cystine values upward and they now account for practically all of the sulfur.

In some cases, the discrepancy between the combined cystine and methionine sulfur and the total sulfur has led to the belief that perhaps another sulfur-containing amino acid exists in certain proteins. The discovery of ergothionine, the betaine of 2-thiol histidine, directed attention to this possibility. It was known that the action of mild oxidizing agents, such as dilute bromine water, on 2-thiol histidine resulted in the removal of the sulfhydryl groups as sulfate. Blumenthal and Clarke<sup>70</sup> tested several proteins with bromine in this manner and obtained from certain of them appreciable quantities of sulfate. Du Vigneaud failed to find any sulfate sulfur in insulin. Until, of course, thiol histidine is unequivocally isolated from a protein digest, its actual occurrence in the protein must remain a matter of conjecture.

The discrepancy alluded to above, between the total sulfur and the combined cystine and methionine sulfur, is very probably due, at least in many cases, and particularly so in the case of insulin, not to a third sulfur-containing amino acid, but to the partial destruction of cystine by severe hydrolytic procedures.

Whether cystine exists in the native protein as such or in whole or part as the reduced form cysteine, remains a problem difficult of solution. Several workers, notably Todrick and Walker<sup>71</sup>, Mirsky<sup>72</sup>, and Greenstein and Edsall<sup>73</sup>, have claimed the demonstration of cysteine as such in native myosin and in the mixed lens proteins. In any case, the protein hydrolysate is usually analyzed in terms of cysteine and the estimate is then given as cystine.

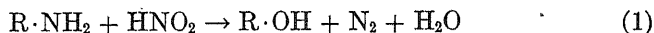
The values of the earlier investigators for both total sulfur and alkali-labile sulfur have recently been confirmed for many proteins. Analyses of the per cent of sulfur, of sulfide sulfur, or of cystine and methionine in certain proteins are recorded in Table 3. These results, calculated as atoms of sulfur or sulfide sulfur or residues of cystine and methionine per gram of protein are also tabulated. It will be noted that for most proteins there is a fair agreement between the number of cystine residues and the expectation on the basis of the sulfide sulfur content, and also between the number of methionine residues and the non sulfide sulfur (total sulfur minus sulfide sulfur) content. The analysis for sulfide sulfur is by no means as reliable as that for total sulfur and so the comparison of the data for the latter and the sum of the cystine and methionine residues is also given. The discrepancy, in the analysis of certain proteins, between alkali-labile sulfur and the sulfur present as cystine or cysteine may, as we have stated, con-



sent another thioamino acid, but, as has already been pointed out, it may also represent destruction of cystine during hydrolysis<sup>11, 74, 75</sup>. Certainly, with few exceptions, alkali-labile sulfur is now accounted for as cystine.

It should be stressed that not all of these analyses are equally reliable, and not all of the proteins equally pure. Thus it is questionable whether the cystine of gelatin represents a part of the protein or an impurity. In the end, studies upon proteins of proven purity with analysis of certain accuracy will supplant or replace many of the results at present available.

**Basic Amino Acids.** Various procedures have been suggested for determining the basic amino acids yielded by proteins on hydrolysis. These substances, histidine, arginine and lysine, are almost completely precipitable by phosphotungstic acid. So isolated, the relative quantities of each have been estimated by Van Slyke in terms of the reaction between their amino groups and nitrous acid. "All the amino acids react with all of their nitrogen, except tryptophane, which reacts with one-half, histidine with one-third, arginine with one-fourth, and proline and oxyproline with none" (reference 76, p. 195). This reaction may be represented as follows:



The  $\alpha$ -amino group reacts far more rapidly than the  $\epsilon$ -amino group of lysine<sup>77</sup>. The rate at which the reaction progresses thus depends, among other properties<sup>78,79</sup>, upon the *same structural conditions that determine* the dissociation of the amino group. The end product is an hydroxy acid, in the case of lysine a di-hydroxy acid.

It should be pointed out that other bases than histidine, arginine and lysine, such as the more recently isolated hydroxylysine, may have been included in these gasometric results.

Methods of estimating the separate bases have been developed in considerable detail. Histidine and arginine form insoluble silver compounds, and Kossel and Kutscher<sup>80</sup> long since devised a method for their isolation which depends upon this behavior. Isolation has certain advantages over indirect estimation. Yields are rarely quantitative, however, and determinations made by isolation methods, though they give minimal results, may be considered satisfactory only in so far as physical chemical studies may be employed to demonstrate or estimate how complete the separation can be, under the conditions that obtain. The precipitation of histidine silver at a neutral reaction and of arginine silver at an alkaline reaction has been shown by Vickery and Leavenworth<sup>81</sup> to depend upon the

<sup>74</sup> Bailey, K., *Biochem. J.*, **31**, 1396 (1937).

<sup>75</sup> Kassel, B., and Brand, E., *Proc. Soc. Exp. Biol. and Med.*, **35**, 444 (1936); *J. Biol. Chem.*, **125**, 435 (1938).

<sup>77</sup> Van Slyke, D. D., *J. Biol. Chem.*, **9**, 185 (1911).

acidity. That is to say, the same structural conditions that determine the dissociation strength of the groups in histidine and arginine influence the solubility of their silver compounds. The determination of the solubility product constants of these silver compounds might not increase the yields that Vickery and Leavenworth have obtained under the conditions found to be optimal for precipitation. They would, however, make the same contribution to these separations as did the determination of the solubility products of the metal sulfides by Bruner and Zawadzki<sup>82</sup>, and permit the estimation of the amounts that can in any case not be separated from a solution which must remain necessarily saturated with respect to the precipitated amino acid compound.\*

The silver distribution method thus depends upon differences in the solubility of the silver salts of these different amino acids. As in the case of all solubility methods, errors may arise either from occlusion of impurities in the precipitates, or from residual solubility of the substances separating from solution. This method should be most accurate for histidine and arginine which precipitate, the one at neutral, the second at strongly alkaline reactions, and least accurate for lysine which remains in solution, and is determined after subsequent precipitation with phosphotungstic acid and crystallization of the picrate.

In addition, specific precipitation methods have been suggested, such as the separation of arginine<sup>83</sup> as flavianate (Kossel<sup>84</sup>), and specific enzymatic methods, such as the arginine method of Hunter and Dauphinee<sup>85</sup>.

Recently an improved method for the determination of arginine as the diflavianate has been developed by Vickery<sup>86</sup> and the results of this method are compared in Table 4 with those from nitrogen distribution and silver precipitation. Comparable improvements in the specific methods for the other two amino acids of this class may well be expected.

Analytical results may be compared on the one hand with another analytical method, or on the other with a physical chemical method. The physical chemical method, depending upon the maximum acid-combining capacity, was suggested as a way of estimating dibasic amino acids over fifteen years ago. (For the earlier literature see Cohn<sup>87</sup>.) The results are included in Table 4, though the method is discussed in Chapter 20.

More recent is the metaphosphoric acid titration of proteins<sup>88, 89</sup>. If, as

<sup>82</sup> Bruner, L., and Zawadzki, J., *Bull. Acad. Sci. Cracovie*, July (1909).

\* The separation of amino acids will not be quantitatively possible until not only the solubility products of their salts but their activity coefficients and the dielectric constants of solutions of amino acids have been estimated. This aspect of the problem is partially considered in a later chapter. The problem of the estimation of the activity coefficients of amino acids has not yet been done to permit calculation of the solvent actions of a mixture of amino acids. See Cohn<sup>87</sup>, and Cohn<sup>88</sup>.

<sup>83</sup> Cohn, E. J., McMeekin, T. L., Ferry, J. D., and Blanchard, M. H., *J. Phys. Chem.*, **43**, 169 (1939).

<sup>84</sup> Kossel, A., *Z. Physiol. Chem.*, **22**, 176 (1896-7).

<sup>85</sup> Hunter, A., *J. Physiol. Chem.*, **22**, 176 (1896-7).

<sup>86</sup> Vickery, L., *J. Biol. Chem.*, **15**, 1 (1935).

TABLE 4. BASIC AMINO ACIDS OF PROTEINS AND ACID COMBINING CAPACITY

	Egg albumin Isolation or Analysis	Insulin Isolation or Analysis	Zein Isolation or Analysis	Gliadin		Casein		Lacto-globulin Isolation or Analysis	Edestin	
				Nitrogen distribution	Isolation or Analysis	Nitrogen distribution	Isolation or Analysis		Nitrogen distribution	Isolation or Analysis
Residues per gram of protein $\times 10^5$										
Histidine	9.8 (1)	69.0 (39)	5.3 (8)	21.6 (11)	21.3 (13)	16.3 (11)	16.1 (15)	9.9 (39)	25.3 (11)	15.6 (39)
Arginine	32.5 (2)	17.5 (39)	9.2 (30)	18.0 (11)	14.8 (2)	20.6 (11)	21.4 (14)	16.6 (39)	90.3 (11)	96.3 (6)
Lysine	34.0 (1)	8.6 (39)	0.0 (8)	4.7 (11)		57.4 (11)	42.8 (15)	66.7 (39)	25.7 (11)	16.2 (39)
Total	76.3	95.1	14.5	44.3		94.3	80.3	93.2	141.3	128.1
Moles of acid bound per gram of protein $\times 10^5$										
	With liquid junction	With liquid junction	With liquid junction	With liquid junction		With liquid junction	Without liquid junction	With liquid junction	With liquid junction	Without liquid junction
	80	101 (6)	17.8 (10)	34 (12)		76 (12)	80 (16)	115 (26)	127 (12)	134 (16)
	87 (3)		21.3 (10)			90				
Moles metaphosphate bound per gram of protein $\times 10^5$										
	78 (4)	103 (7)								137 (7)
	78	110								138 (7)
	80									

The values for histidine, arginine and lysine of most proteins were tabulated in 1931 by Cohn (Table 20, Ref. 28).

**Egg albumin.** Values of Vickery and of Vickery and Shore have been reported. The earlier values of Osborne, Jones and Leavenworth (29) of  $11.00 \times 10^{-5}$  for histidine,  $28.2 \times 10^{-5}$  for arginine and  $25.7 \times 10^{-5}$  for lysine give only a total of 64.9 residues  $\times 10^{-5}$  per gram of egg albumin.

**Zein.** Vickery (30) redetermined the bases and reported estimates of 0.77% for histidine, in close agreement with the value of Osborne and Liddle (8), 0.82%, and of 1.6% for arginine, which is in good agreement with that of Osborne and Liddle, but lower than that of Kossel and Kutscher (9), 1.82%.

**Gliadin.** The histidine analysis of Van Slyke (11), 3.35%, has now been confirmed by Haugaard and Johnson (13) (Tables 46 and 37). Recalculated from nitrogen distribution to per cent amino acid in the protein, their results on histidine range only from 3.21% to 3.31%. On the basis of this determination molecules of molecular weight 26,000 as revealed by the ultracentrifuge would contain 6 histidine molecules and  $4\frac{1}{2}$  cystine residues, or if the molecular weight be double the minimum molecular weight previously suggested (31) as indicated by the osmotic pressure measurements (32), there will be 4 cystine and 9 histidine residues in the molecule. The histidine contents of the various gliadin preparations examined might be expected to be closely the same and thus to preclude a molecular weight of 20,700 though not of 41,000.

Haugaard and Johnson (13) report a larger value for arginine, namely, 3.74%, than Vickery.

**Casein.** The histidine and arginine results of Osborne and Guest (15) and of Van Slyke (33) are in fair agreement though lower than the values of Hoffman and Gortner (34) on casein and of Sandstrom (35) on deaminized casein (see Table 20, Ref. 28).

Vickery and White's (36) results are in excellent agreement for arginine with those of Osborne and Guest (15), but are considerably lower than those of Osborne and Guest and Van Slyke in the case of histidine, since they report as an average 1.83% as against 2.50% and 2.53% of the earlier workers. Their value for lysine is 1.1%

TABLE 4. *Continued*

	Gelatin		Hemoglobin (horse)		Limulus hemocyanin		Myosin (rabbit)	Pseudo- globulin	Fibrin (cattle)		Serum albumin (horse)
	Nitrogen distribu- tion	Isola- tion or Analysis	Nitrogen distribu- tion	Isolation or Analysis	Nitrogen distribu- tion	Isolation or Analysis	Isolation or Analysis	Isolation or Analysis	Nitrogen distribu- tion	Isolation or Analysis	Nitrogen distribu- tion
Residues per gram of protein $\times 10^5$											
Histidine ...	18.9 (11)		49.9 (18)	49.3 (19)	54.5 (11)	29.1 (25)	11.2 (20)		19.2 (11)	16.1 (27)	21.9 (24)
Arginine....	47.9 (11)	49.9 (26)	23.6 (18)	20.6 (2)	48.6 (11)	36.6 (25)	40.4 (20)		41.9 (11)	44.2 (2,27)	28.1 (24)
Lysine.....	40.5 (11)		65.4 (18)	55.4 (19)	52.4 (11)	61.1 (25)	68.0 (20)		69.4 (11)		90.3 (24)
Total .....	107.3		138.9	125.3	155.5	126.8	119.6		130.5		140.3
Moles of acid bound per gram of protein $\times 10^5$											
	With liquid junction	Without liquid junction	With liquid junction				With liquid junction	With liquid junction			With liquid junction
	89 (12)	96 (16)	148 (18)				155 (21)	92			140 (23)
	92 (12)		(134) <sup>a</sup>								
Moles metaphosphate bound per gram of protein $\times 10^5$											
			149 (7)				150 (7)	90 (7)			138 (7)
			150 (7)								141
											143

higher than that of Osborne and Guest. The sums of the bases reported by Osborne and Guest, 11.76%, and by Vickery and White, 11.93%, are in excellent agreement.

Van Slyke's value for the lysine content of casein is, however, far larger than that reported by Osborne and Guest, 5.59%, by Hoffman and Gortner (34), 6.91%, by Vickery and White, 6.3%, or by Block, Jones and Gersdorff (37) who obtain results of 6.5 and 6.6% by isolation and conclude that the higher results given by the Van Slyke method depend upon the presence of a lysine precursor in casein. On the basis of a molecular weight of 96,000 there are presumably 16 histidine, 21 arginine and 42 lysine residues.

**Lactoglobulin.** Chibnall's (39) recent data given in the table differ but slightly from Cannan's (26) for arginine and lysine, but the histidine value is appreciably lower.

**Edestin.** The basic amino acids of edestin have been determined repeatedly. Vickery and Leavenworth (17) isolated 2.19% of lysine, whereas Van Slyke (11) and Sandstrom (35) using the nitrogen distribution method estimated 3.76 and 3.51% respectively. The agreement in the case of histidine is by no means so satisfactory. Vickery and Leavenworth (17) isolated 2.08%, Van Slyke (11) estimated 3.92%, Sandstrom (35) 2.85%, and Hanke and Kossler (38) 3.04%.

**Gelatin.** Lower values for arginine, namely, 8.22% or  $47 \times 10^{-5}$  groups per gram, were employed in our earliest calculations (12). The total bases on this computation were therefore 106.6 free basic and 41.4 free acid groups. The sum, 148, is essentially identical with that in the above calculation, namely, 146.5, which follows from adopting analytical figures selected by Bergmann and Niemann (27).

**Hemoglobin.** Abderhalden (22) made some early analyses upon horse hemoglobin. Whereas the sum of the bases that he reported, 19.8%, is not very different from the above sum of histidine, arginine and lysine, 19.33%, the distribution of bases is very different.

**Fibrin.** Bergmann and Niemann's (27) new analyses confirm the magnitude of Van Slyke's estimate of the hexone bases. Vickery's (2) analysis of arginine by the

TABLE 4. Continued

(1)	Vickery, H. R.	<i>Biochem. J.</i> , <b>26</b> , 1101 (1932).
(2)	Vickery, H. R.	<i>Biochem. J.</i> , <b>26</b> , 1101 (1932).
(3)	Cohn, J.	<i>Biochem. J.</i> , <b>6</b> , 10 (1938).
(4)	Perlmann, T.	<i>Biochem. J.</i> , <b>32</b> , 926 (1938).
(5)	du Vigneau, J.	<i>Bor Symp. Quant. Biol.</i> , <b>6</b> , 275 (1938).
(6)	Harington, J. A.	<i>Biochem. J.</i> , <b>30</b> , 809 (1936).
(7)	Perlmann, T.	<i>Biochem. J.</i> , <b>26</b> , 995 (1910).
(8)	Cohn, J.	<i>Biochem. J.</i> , <b>30</b> , 809 (1936).
(9)	Cohn, J.	<i>Biochem. J.</i> , <b>30</b> , 809 (1936).
(10)	Cohn, J.	<i>Biochem. J.</i> , <b>30</b> , 809 (1936).
(11)	Van Slyke, J.	<i>Biochem. J.</i> , <b>105</b> , 319 (1934).
(12)	Cohn, E. J.	<i>d. biol. Arbeitsmeth.</i> , Abt. 1, Part 7, 51 (1923).
(13)	Haugaard, E.	<i>Carlsberg</i> , <b>18</b> , No. 2, (1930).
(14)	Vickery, H. R.	<i>Biochem. J.</i> , <b>3</b> , 425 (1911).
(15)	Cohn, J.	<i>Bor Symp. Quant. Biol.</i> , <b>6</b> , 24 (1938).
(16)	Cohn, J.	<i>C. S., J. Biol. Chem.</i> , <b>76</b> , 707 (1928).
(17)	Cohn, J.	<i>C. S., J. Biol. Chem.</i> , <b>76</b> , 707 (1928).
(18)	Richard, M. H.	<i>C. S., J. Biol. Chem.</i> , <b>54</b> , 509 (1937).
(19)	Cohn, J.	<i>C. S., J. Biol. Chem.</i> , <b>54</b> , 509 (1937).
(20)	Cohn, J.	<i>C. S., J. Biol. Chem.</i> , <b>54</b> , 509 (1937).
(21)	Cohn, J.	<i>C. S., J. Biol. Chem.</i> , <b>54</b> , 509 (1937).
(22)	Cohn, J.	<i>C. S., J. Biol. Chem.</i> , <b>54</b> , 509 (1937).
(23)	Cohn, J.	<i>C. S., J. Biol. Chem.</i> , <b>54</b> , 509 (1937).
(24)	Cohn, J.	<i>C. S., J. Biol. Chem.</i> , <b>54</b> , 509 (1937).
(25)	Cohn, J.	<i>C. S., J. Biol. Chem.</i> , <b>54</b> , 509 (1937).
(26)	Cannan, R. K., Palmer, A. H., and Kibrick, A.	<i>J. Biol. Chem.</i> , <b>142</b> , 803 (1942).
(27)	Richard, M. H.	<i>C. S., J. Biol. Chem.</i> , <b>115</b> , 77 (1936).
(28)	Richard, M. H.	<i>C. S., J. Biol. Chem.</i> , <b>115</b> , 77 (1936).
(29)	Richard, M. H.	<i>C. S., J. Biol. Chem.</i> , <b>115</b> , 77 (1936).
(30)	Richard, M. H.	<i>C. S., J. Biol. Chem.</i> , <b>115</b> , 77 (1936).
(31)	Richard, M. H.	<i>C. S., J. Biol. Chem.</i> , <b>115</b> , 77 (1936).
(32)	Richard, M. H.	<i>C. S., J. Biol. Chem.</i> , <b>115</b> , 77 (1936).
(33)	Richard, M. H.	<i>C. S., J. Biol. Chem.</i> , <b>115</b> , 77 (1936).
(34)	Richard, M. H.	<i>C. S., J. Biol. Chem.</i> , <b>115</b> , 77 (1936).
(35)	Richard, M. H.	<i>C. S., J. Biol. Chem.</i> , <b>115</b> , 77 (1936).
(36)	Richard, M. H.	<i>C. S., J. Biol. Chem.</i> , <b>115</b> , 77 (1936).
(37)	Richard, M. H.	<i>C. S., J. Biol. Chem.</i> , <b>115</b> , 77 (1936).
(38)	Richard, M. H.	<i>C. S., J. Biol. Chem.</i> , <b>115</b> , 77 (1936).
(39)	Richard, M. H.	<i>C. S., J. Biol. Chem.</i> , <b>115</b> , 77 (1936).

is suggested, combination of the metaphosphoric acid takes place only with those groups of the proteins which are positively charged, then the sum of the bases should be given by a simple phosphorus determination (correcting of course for the phosphorus of the untitrated protein) of the precipitate on which this method depends.

**Tryptophane and Tyrosine.** Whereas tryptophane and tyrosine are present in proteins to only a small extent, the colorimetric methods that have been developed for their estimation appear to be reliable at least as a first approximation. The percentage of these two amino acids estimated to be present in a number of proteins are tabulated in Table 5. The earlier literature, including the work of Folin and his collaborators and of Linderstrøm-Lang on casein, has previously been summarized (reference 90, Table 19). Not included in this summary are studies on *egg albumin* by Calvery<sup>91</sup> and on *insulin* by du Vigneaud<sup>92</sup>.

On the basis of the tryptophane estimates of a number of investigators, the molecular weight of *zein* would have to be many times that observed either by osmotic pressure or ultracentrifugal measurements.

The fractions of *gliadin* separated by Haugaard and Johnson<sup>93</sup>, although

Van Slyke<sup>94</sup>, vary appreciably in the percentage of some of the other amino acids studied. Their least soluble fraction, IV, had a tryptophane content of 1.08%, in excellent agreement with a number of previous determinations (reference 90, p. 861). The tryptophane contents of their other fractions were higher than of their fraction IV, and increased with the solubility and the acid-insoluble humin nitrogen, reaching a value of 1.82%. Conversely, the tyrosine content of their least soluble fraction was highest, 3.02%, and close to that given by Looney<sup>95</sup>, 3.04%.

The tryptophane content of *casein* has repeatedly been studied; most results fall between 1.4 and 1.6%, the higher value being due to Linderstrøm-Lang<sup>96</sup>. The values for tyrosine reported are very discrepant. The tryptophane results reported for *edestin* generally range from 1.45 to 1.56% and for tyrosine from 4.53 to 4.58%<sup>90</sup>.

The tryptophane content reported by early workers for *serum albumin* was always high, perhaps because of the difficulties in purification, since serum albumin can be separated into a number of crystalline fractions of different solubility<sup>97</sup>, different dipole moment<sup>98</sup>, different carbohydrate content<sup>97, 99</sup>, and different tryptophane content. Fürth and Lieben<sup>100</sup> reported 1.3%, and Hunter and Borsook<sup>101</sup> 1.79% of tryptophane recovered from serum albumin. Folin and Marenzi<sup>102</sup> studied a preparation of serum albumin that had been recrystallized from five to seven times in this laboratory and found 0.52 and 0.53% of tryptophane in the fractions studied. More recently Hewitt<sup>99</sup> had studied a serum albumin recrystallized twelve times and obtained a value just half of that reported by Folin and Marenzi<sup>102</sup>, namely, 0.26%, whereas McMeekin<sup>103</sup>, on a carbohydrate-free crystalalbumin preparation, recrystallized as the albumin sulfate, has confirmed the earlier result of Folin and Marenzi. Although tryptophane appears to differ in the various fractions of serum albumin, the tyrosine content seems to be relatively constant. Hunter and Borsook<sup>101</sup> reported 4.63%, Folin and Marenzi<sup>102</sup> 4.66 and 4.67%, and Hewitt's<sup>99</sup> crystalalbumin preparation yielded 4.79%.

Sulfur and sulfide sulfur, tryptophane, tyrosine and cystine appeared to be the most accurately estimated constituents of the protein molecule which were available some years since<sup>54, 104</sup>. In so far as other analytical procedures become sufficiently accurate and protein preparations purer, the newer results should supplement and render ever more accurate those reported in Table 5.

<sup>94</sup> Van Slyke, D. D., *Handb. d. biol. Arbeitsmeth.*, Abt. I, Part 7, 51 (1923).

<sup>95</sup> Looney, J. M., *J. Biol. Chem.*, 69, 519 (1926).

<sup>96</sup> Linderstrøm-Lang, K., *Compt. rend. trav. lab. Carlsberg*, 17, No. 9, 48 (1929).

<sup>97</sup> Sørensen, S. P. L., *Compt. rend. trav. lab. Carlsberg*, 18, No. 5, 1 (1930).

<sup>98</sup> ...

TABLE 5. THE AMINO ACID COMPOSITION OF THE PROTEINS

Insulin %	Zein %	Gliadin %	Casein %	Edestin %	Coconut globulin %	Gelatin %	Hemo- globin (horse) %	Serum albumin %	Silk fibroin %	Spider silk fibroin %	Wool keratin %	Salmin %	Fibrin (cattle) %
0.0 (74)	0.0 (18)		0.45 (27)	3.80 (34)	Trace (39)	25.5 (41)			43.8 (50)	35.13 (54)	0.6 (55, 56)		
3.57 (73)	9.79 (12)	2.0 (19)	1.85 (27)	3.60 (34)	4.11 (39)	8.70 (41)		2.08 (45)	26.4 (50)	23.40 (54)	4.4 (55, 56)		
2.66 (73)	1.02 (12)	0.13 (19)	5.0 (28)	0.33 (34)	1.76 (39)	3.3 (28)		0.60 (45)	13.87 (51)		2.9 (55, 56)	7.80 (61)	
			3.5 (29)			1.4 (29)			1.36 (51)				
	1.88 (12)	3.34 (18)	7.93 (27)	5.6 (35)	3.57 (39)	0.0 (41)					2.8 (55, 56)	4.30 (61)	
30.00 (8)	25.00 (13)	6.62 (18)	9.70 (27)	20.90 (34)	5.96 (39)	7.10 (41)		20.0 (45)	2.50 (52)	0.76 (54)	11.5 (55, 56)		
	9.04 (12)	13.22 (18)	8.70 (30)	4.10 (12)	5.54 (39)	19.7 (42)		1.04 (45)	1.00 (52)	3.68 (54)	4.4 (55, 56)	11.00 (61)	5.1 (42)
	0.8 (75)		0.23 (18)	2.00 (34)	2.05 (39)	14.4 (42)							
	7.6 (13)	2.35 (19)	3.88 (27)	3.09 (12)				3.08 (45)	1.50 (52)				
	2.35 (14)	2.03 (14)	3.25 (14)	2.39 (36)		1.40 (41)			2.59 (56)		0.67 (57)		2.59 (36)
12.50 (9)	0.91 (15)	2.40 (20)	0.42 (31)	1.36 (20)	1.46 (39)	0.17 (21)	1.02 (31)	5.71 (46)					1.51 (15)
		1.10 (21)	1.54 (32)	1.46 (4)	1.25 (39)	0.0 (32)	1.28 (4)	0.53 (4, 47)			1.8 (56)		3.0 (62)
12.5 (74)	5.90 (16)	3.04 (21)	5.36 (32)	4.54 (4)	3.18 (39)	0.0 (32)	3.15 (4)	4.79 (48)	13.2 (50)	8.20 (54)	4.8 (56, 58)		6.5 (32)
10.7 (33)	6.82 (12)	3.30 (22)	2.50 (18)	2.41 (33)	2.42 (40)	2.94 (24)	7.64 (43)	3.40 (49)	0.67 (53)		0.66 (59)		2.50 (42)
3.05 (33)	1.60 (6, 17)	2.57 (6, 23)	3.72 (6)	16.76 (6)	15.92 (40)	8.68 (6)	3.59 (6)	4.90 (49)	0.76 (6)	5.24 (54)	10.2 (56)	87.40 (61)	7.70 (6, 42)
1.26 (33)	0.0 (12)	0.69 (24)	6.25 (18)	2.37 (33)	5.80 (40)	5.92 (24)	8.10 (43)	13.20 (49)	0.25 (53)		2.8 (56)		10.14 (24)
30.00 (11)	1.8 (13)	0.77 (25)	5.95 (33)	12.0 (33)	5.12 (39)	3.40 (41)	8.9 (44)	3.12 (45)		11.70 (54)	7.27 (60)		5.9 (42)
	31.30 (13)	43.66 (18)	21.6 (30)	20.7 (33)	19.07 (39)	5.80 (41)	6.3 (44)	1.82 (45)			15.27 (60)		14.1 (42)
	2.50 (13)	7.70 (26)	10.5 (30)		0.0 (39)	0.0 (41)							
1.65 (33)	3.64 (12)	5.22 (18)	1.61 (18)	2.15 (33)	1.57 (39)	0.40 (41)	1.22 (44)	1.29 (49)		1.16 (54)	1.37 (60)		
107.89	105.95	100.14	103.94	109.56	78.78	108.81	41.20	65.86	107.00	89.27	71.44	110.50	59.04

ies recorded in this table have previously been tabulated and discussed by Cohn (63) and Vickery (64). In 1922 two estimates of the tryptophane in egg albumin were published, one of 1.11% by May and Rose (65) and one of 1.23% by Folin and Marenzi (4). A large number of analyses have since been reported of which the Folin and Marenzi value (4), 1.16, may be considered the lowest and the (16) value of 1.30 the highest. The redetermination by Calvery (2) is 1.28%. A new unpublished determination by Mehl and McMeekin (133%) Vickery (64) accepts the value of Folin and Marenzi, which is therefore tentatively given here, namely, 1.18%.

tyrosine are at least as consistent as those of tryptophane. Folin and Looney (32) estimated that egg albumin contains 4.20% tyrosine in 2) in 1921 estimated that there was 4.21% present. In the intervening years Looney (21) reported an estimate of 4.1, Folin and Ciocalteu in and Marenzi (4) values of 4.03 and 3.93%. Calvery's value (2) for aspartic acid is 8.00 and for glutamic acid 13.96%, which are lower in the table. The value for ammonia, 1.18%, of Shore and his coworkers (7) is close to Chibnall's data.

t values (33) for the bases are: arginine 5.63, histidine 1.45, and lysine 5.06%. These closely confirm the earlier data given in the table. new values for tyrosine (4.10%) and tryptophane (1.32%), likewise in close agreement with the literature.

stidine value of Chibnall (33) is distinctly higher and the lysine value distinctly lower than the earlier figures given by du Vigneaud ) and 2.26% respectively. The tyrosine value given in the table is slightly higher than the value of 12.2% given in ref. 10. No value ms to be available.

Ciocalteu's value for tyrosine is tentatively accepted for zein, though it is higher than that of Folin and Denis (66), Folin and Looney Marenzi (4). No more than 0.2% of tryptophane has ever been reported for zein and a large number of investigators believe that zein hane. (See Table 19, Ref. 63).

ryptophane and tyrosine values given are those of Looney (21), which came closest to the least soluble fraction IV studied by Haugaard ue assumed for tryptophane is that of Folin and Looney and the mean of that of Linderström-Lang (67). The tyrosine value of Linder- ations were lower even than those of Looney, 4.5% (21), of Folin and Looney (32), 5.36%, and of Folin and Ciocalteu (16), 6.55%. Clearly, of carefully fractionated casein will have to be redetermined. The intermediate value of Folin and Looney is given to indicate that amount of tyrosine is present in this protein. The recent and highly accurate method of serine determination with periodate developed in (28), gives far higher values for serine than earlier determinations (18). These authors have also determined threonine values for y have concluded from their analyses for hydroxyamino acids that "hydroxyglutamic acid does not occur, to any considerable extent, if 8, p' 691).

ryptophane content of edestin has been repeatedly studied. Most of the results lie between 1.40 and 1.56 (32, 16, 4, 21) with a mean near if. 63). This average value is in excellent agreement with Bailey's determination (20). In the case of tyrosine, however, later analyses lue, close to 4.55 (63, Table 19), and Bailey's determinations are still lower and range from 4.10 to 4.24%. Chibnall's recent values (33) sine are appreciably higher than the earlier values of Vickery and Leavenworth (37), namely 2.08 and 2.19% respectively.

ptophane and tyrosine seem to be present in gelatin.

0.53% has been tentatively accepted for tryptophane, though the earlier values of Fürth and Lieben (68), 1.30%, and of Hunter and Bor- much higher. Serum albumin can be separated into a number of crystalline fractions of different solubility (70), different dipole moment hydrate content (70, 48) and different tryptophane content. Thus the tryptophane content reported by early workers was always high, the difficulties in purification since serum albumin is rich in this amino acid. Folin and Marenzi (4) studied a preparation of serum al- n recrystallized from 5 to 7 times in this laboratory and found 0.52 and 0.53% of tryptophane in the fractions studied. More recently tied a serum albumin recrystallized 12 times and obtained a value just half of that reported by Folin and Marenzi, namely, 0.26%.

phane appears to be a variable in the various fractions of serum albumin, the tyrosine content would appear to be invariant. Hunter ported 4.63%, Folin and Marenzi (4) 4.66% and 4.67%, and Hewitt's crystalalbumin preparation yielded 4.79% (48).

discrepancy inheres in the various values reported for tryptophane. Bergmann and Niemann (42) adopt the value of Holm and Green- ich is nearly as high as the value reported by Fürth and Lieben (68) and far higher than the estimates of Folin and Looney (32) and of d Moeller (72).



TABLE 5. Continued

- nes, D. B., and Leavenworth, C. S., *Am. J. Physiol.*, 24, 252 (1909).  
*Biol. Chem.*, 94, 613 (1931).  
 and, E., *J. Biol. Chem.*, 125, 145 (1938).  
 enzi, A. D., *J. Biol. Chem.*, 83, 89 (1929).  
 d Shore, A., *Biochem. J.*, 26, 1101 (1932).  
*Biol. Chem.*, 132, 335 (1940).  
 H., and Stueck, G., *J. Biol. Chem.*, 112, 407 (1936).  
 intersteiner, O., *J. Biol. Chem.*, 98, 1281 (1932).  
 Vigneaud, V., *J. Biol. Chem.*, 116, 101 (1937).  
 lensen, H., and Wintersteiner, O., *J. Pharm. and Expt'l. Therapy*, 32, 367 (1928).  
*Cold Spring Harbor Symp. Quant. Biol.*, 6, 275 (1938).  
 d Liddle, L. M., *Am. J. Physiol.*, 26, 286 (1910).  
*Physiol. Chem.*, 130, 59 (1923) (1923); 115, 25 (1935).  
 J., *Biol. Chem.*, 91, 681, 689 (1923).  
 id White, A., *Biol. Chem.*, 99, 701 (1933).  
 caleu, V. J., *Biol. Chem.*, 73, 627 (1927).  
*impl. rend. trav. lab. Carlsberg*, 22, 519 (1938).  
 id Cluett, H. H., *J. Biol. Chem.*, 9, 333, 425 (1911).  
 id Clapp, S. H., *J. Physiol.*, 17, 231 (1906).  
*Biol. Chem.*, 69, 519 (1926).  
 id Johnson, E., *Compt. rend. trav. lab. Carlsberg*, 18, No. 2 (1930).  
 id Campbell, B. D., *Biol. Chem.*, 63, 225 (1925).  
*Hand. biol. Arithmet.*, Abt. I, Teil 7, 51 (1923).  
 Moeller, O., *J. Biol. Chem.*, 79, 439 (1923).  
 Wilson, R., *Cereal Chem.*, 5, 473 (1928).  
 d Shinn, L. A., *J. Biol. Chem.*, 139, 687 (1941).  
 d Shinn, L. A., *J. Biol. Chem.*, 138, 91 (1941).  
 id Nicolet, B. H., *J. Biol. Chem.*, 12, 290 (1918).  
 id L., and Quackenbush, F. W., *Ber.*, 72, 407 (1939).  
 ney, J. M., *J. Biol. Chem.*, 51, 421 (1922).  
*Merikan Lecture* (1942), *Proc. Roy. Soc. (London)*, Ser. B (1943), in press.  
*Z. Physiol. Chem.*, 37, 499 (1902).  
 id Van Slyke, D. D., *J. Biol. Chem.*, 6, 419 (1909).  
*J. Biol. Chem.*, 106, 451 (1934).  
 id Leavenworth, C. S., *J. Biol. Chem.*, 76, 797 (1928).  
 (38) Osborne, T. B., and Harris, I. F., *J. Am. Chem. Soc.*, 25, 323 (1903).  
 (39) Jones, D. B., and Johns, C. O., *J. Biol. Chem.*, 44, 291 (1920).  
 (40) Johns, C. O., Finks, A. J., and Gersdorff, C. E., *J. Biol. Chem.*, 37, 149 (1919).  
 (41) Dakin, H. D., *J. Biol. Chem.*, 44, 498 (1920).  
 (42) Bergmann, M., and Niemann, C., *J. Biol. Chem.*, 115, 77 (1936).  
 (43) Vickery, H. B., and Leavenworth, C. S., *J. Biol. Chem.*, 79, 377 (1928).  
 (44) Chibnall, A. C., and Leavelle, C., *Unpublished work*.  
 (45) Abderhalden, E., *Z. Physiol. Chem.*, 17, 84 (1902).  
 (46) Sullivan, M. X., *Pub. Health Rep.*, U. S. P. H. S. Suppl. No. 86 (1930).  
 (47) Harkins, T. B., *Food Res.*, 21, 2884 (1933).  
 (48) Hewitt, F. B., *Biochem. J.*, 8, 541 (1914).  
 (49) Harder, P., *Biochem. J.*, 30, 229 (1936); *ibid.*, 31, 360 (1937).  
 (50) Eardley, M., and Niemann, C., *J. Biol. Chem.*, 122, 577 (1938).  
 (51) Nicolet, B. H., and Shinn, L. A., *J. Biol. Chem.*, 140, 685 (1941).  
 (52) Abderhalden, E., *Z. Physiol. Chem.*, 120, 207 (1922).  
 (53) Vickery, H. B., and Block, R. J., *J. Biol. Chem.*, 93, 105 (1931).  
 (54) Fischer, E., *Z. Physiol. Chem.*, 53, 126 (1907).  
 (55) Abderhalden, E., and Votinovici, A., *Z. Physiol. Chem.*, 52, 368 (1907).  
 (56) Marston, H. R., *Compt. of Sci. and Ind'l. Res., Com'lth of Australia, Bull.*, 38 (1938).  
 (57) Barritt, J., *Biochem. J.*, 28, 1 (1934).  
 (58) Whewell, C. S., Ph.D., Thesis, University of Leeds (1935).  
 (59) Vickery, H. B., and Block, R. J., *J. Biol. Chem.*, 86, 107 (1930).  
 (60) Speakman, J. B., and Townsend, F., *Nature*, 139, 411 (1937).  
 (61) Kossel, A., and Dakin, H. D., *Z. Physiol. Chem.*, 41, 407 (1904).  
 (62) Holm, G. E., and Greenbank, C. R., *J. Am. Chem. Soc.*, 45, 1788 (1923).  
 (63) Cohn, E. J., *Ergebn. d. Physiol.*, 33, 781 (1931).  
 (64) Vickery, H. B., *Annals New York Acad. Sci.*, 41, 87 (1941).  
 (65) May, C. E., and Rose, E. R., *J. Biol. Chem.*, 54, 213 (1922).  
 (66) Folin, O., and Denis, W., *J. Biol. Chem.*, 12, 245 (1912).  
 (67) Linderström-Lang, K., *Compt. rend. trav. lab. Carlsberg*, 17, No. 9, 48 (1929).  
 (68) Furth, O., and Lieben, F., *Biochem. Z.*, 109, 124 (1920).  
 (69) Hunter, A., and Borsook, H., *J. Biol. Chem.*, 57, 507 (1923).  
 (70) Sørensen, S. P. L., *Compt. rend. trav. lab. Carlsberg*, 18, No. 5 (1930).  
 (71) Ferry, J. D., and Onley, J. L., *J. Am. Chem. Soc.*, 60, 1123 (1938).  
 (72) Jones, D. B., Gersdorff, C. E. F., and Moeller, O., *J. Biol. Chem.*, 62, 183 (1924).  
 (73) Nicolet, B. H., and Shinn, L. A., *J. Am. Chem. Soc.*, 63, 1486 (1941).  
 (74) Harington, C. R., and Neuberger, A., *Biochem. J.*, 30, 809 (1936).  
 (75) Furth, O., and Minnebeck, H., *Biochem. Z.*, 250, 18 (1932).

**Aliphatic Amino Acids.** Amino acids and their derivatives are most soluble in alcohols when conditions render them most nearly like non-electrolytes. Several procedures in the separation of amino acids from one another depend upon this phenomenon. Thus Dakin<sup>105, 106, 107</sup> discovered that the monoamino-monocarboxylic acids could be extracted by means of higher alcohols from neutral aqueous solutions. In practice the neutralized mixtures of amino acids from a protein hydrolysate were continuously extracted in a Kutscher apparatus by means of butyl alcohol. The monoamino-monocarboxylic acids were extracted by this solvent, although the solubility in it of most of them is very small (see<sup>108</sup>). Through continuous distillation, preferably under reduced pressure, fresh portions of butyl alcohol perfuse the aqueous solution until extraction is complete. In this process monoamino-monocarboxylic acids were separated from the basic and dicarboxylic amino acids. The electrical properties of tyrosine, though it is trivalent, do not prevent its extraction by butyl alcohol, since it is neutral near pH 6. It is also relatively insoluble at this reaction, and is generally removed before extraction. As further evidence that the extraction of amino acids depends upon electrical conditions, it may be mentioned that glutamic and aspartic acids are largely extractable at pH 3, and histidine near pH 7.5.\*

Proline alone among the amino acids that are extracted by butyl alcohol is relatively soluble in butyl alcohol. Because of the solubility of proline in alcohols, the name prolamines has been given to proteins which are alcohol soluble, and which generally contain proline in large amounts. It is probable that the solubility of this class of substances does not depend entirely, however, on the presence of this amino acid since there are proteins rich in proline which are not alcohol soluble.

The monoamino-monocarboxylic acids other than proline, though they are extractable, are by no means soluble in the higher alcohols<sup>108</sup>. It is necessary to assume that the undissociated molecule, or one of the ionic species with which it is in equilibrium, is soluble to some extent in alcohol, or in the alcohol-water mixture.† It seems probable that the *neutral molecules* rather than the ions are extracted. Dakin has noted that diketopiperazines are readily extracted by butyl alcohol, and Simms<sup>109</sup> has shown that alanyl-alanine anhydride shows no titration constants throughout the neutral range. The diketopiperazines bear an intermediate position between proline and the monoamino-monocarboxylic acids that are extracted by butyl alcohol. The latter are insoluble in ethyl alcohol, the diketopi-

<sup>105</sup> Dakin, H. D., *Biochem. J.*, **12**, 290 (1918).

<sup>106</sup> Dakin, H. D., *J. Biol. Chem.*, **44**, 499 (1920).

<sup>107</sup> Dakin, H. D., *J. Biol. Chem.*, **44**, 499 (1920).

perazines slightly soluble and proline more soluble. Here again it is probable that a balance between the non-polar properties of groups and the electrical properties of the molecule determine the order of solubility of otherwise comparable substances in alcohol-water mixtures.

The butyl alcohol extraction method may be employed in a partial separation of the aliphatic monoamino-monocarboxylic amino acids, but the problem of interaction complicates this procedure and the presence of large amounts of one amino acid may, as Dakin has pointed out, interfere with the quantitative extraction of another.

The separation of the aliphatic  $\alpha$ -amino acids from protein hydrolysates and from each other has in the past largely been carried out by Fischer's method of distilling their methyl or ethyl esters. The results led to our first knowledge of the nature and distribution of these amino acids present in most proteins. The amino acids are, as we have seen, such strongly polar molecules that their melting points, though lower than those of inorganic salts, are higher than those of most organic compounds. The melting points of amino acids are not accurately known since most decompose when melting, but are in the neighborhood of  $300^{\circ}$ , reflecting dipole ionic structure as Pfeiffer<sup>110</sup> and Bjerrum<sup>111</sup> suggested.

Many of the amino acids sublime at temperatures considerably lower than their melting points. The temperatures of sublimation of most of the amino acids have recently been reported, and also the melting points of the sublimates<sup>112</sup>. The latter prove to be very close to the melting points of the amino acids as carried out by improved methods<sup>113, 114</sup>.

Dunn and Brophy<sup>114</sup> give the melting point of glycine as  $289$  to  $292^{\circ}$ , of di- and triglycine as  $262$  to  $265^{\circ}$ . Methyl hydantoic acid, which has the same composition as glycylglycine<sup>115</sup> and which is presumably not a dipolar ion, melts almost  $100^{\circ}$  lower, at  $170^{\circ}$ . The melting points of the hydantoic acids are in the same range as those of the sugars; they are higher than that of urea,  $133^{\circ}$ , and somewhat lower than that of biuret,  $193^{\circ}\text{C.}$ , acetyl urea,  $215^{\circ}$ , or carbonyl diurea,  $233^{\circ}$ .

The melting points of the amino acids are thus far too high, as are their sublimation points, to permit their distillation without decomposition, whereas some of their derivatives can be so distilled. Emil Fischer<sup>116</sup> converted the amino acids into their ethyl or methyl esters which could be distilled. Since those with the larger paraffin side chains distilled at the higher temperatures, partial separations could be effected by fractional distillation. The method yielded many of the available results, but pre-

<sup>110</sup> Pfeiffer, P., *Berichte*, 55, 1762 (1922).

sumably because of the polar groups of amino acid esters, the increment in boiling point for each  $\text{CH}_2$  group is far smaller than for the simple paraffins.

The use of esters of still higher alcohols has been considered, but whereas the proportion of non-polar groups would be thereby increased, so would the absolute boiling point. The use of the acetylated amino acid esters has been introduced by Chéribuliez<sup>117</sup>. The acetylated esters are considerably more stable than the simple esters of the amino acids and by partial alkaline hydrolysis the ester group may be removed. The acetyl amino acids so obtained may be crystallized in pure form more readily than the amino acids themselves<sup>118</sup>. Separation of pure individuals by fractional distillation, however, seems to be no more easily carried out with the acetylated esters than by the Fischer technique. It is thus presumably the polar groups which reduce the spread in the boiling points and render more difficult the development of fractional distillation methods of comparable effectiveness to those developed for hydrocarbons of different chain lengths.

"Glycine, alanine, proline, oxyproline, leucine, and phenylalanine until recent years could be separated only by the Fischer procedure and, although much valuable data was secured, the results can only, even under the best conditions and with the most careful attention to details, have been approximations. There is no way in which one can calculate what the error in these determinations may have been. Doubtless many of the values in the literature approach the truth quite closely; others, however, must be much too low and some may even be too high" (reference 14, p. 105).

The search for specific precipitants for the amino acids is now being successfully prosecuted by Bergmann and his co-workers. Moreover, these investigators are correcting for the residual solubility of the precipitated compounds. As a result we already have a few accurate analyses of glycine and of proline and of the other amino acids of this class and may confidently expect reliable estimates both of other amino acids in these proteins and of these amino acids in other proteins.

"The method for glycine<sup>57, 119</sup> involves the precipitation of the complex compound of potassium trioxalato chromiate,  $[\text{Cr}(\text{C}_2\text{O}_4)_3]\text{K}_3 + 3\text{H}_2\text{O}$ , that is formed when the reagent is added, together with two volumes of alcohol, to the solution that contains glycine. Under carefully controlled conditions of temperature, acidity and time, a constant proportion of the glycine (about 88 per cent) is precipitated. With the use of this value, the quantities of glycine precipitated from a protein hydrolysate are corrected and an estimate is made of the glycine yielded

"The determination of alanine requires the previous removal of glycine and involves a series of small scale determinations of the nitrogen precipitated by dioxalatodipyridinochromiato acid (dioxopyridic acid),  $[\text{Cr}(\text{C}_2\text{O}_4)_2 \cdot (\text{C}_5\text{H}_5\text{N})_2] \cdot \text{H}^{120}$ , in order to find the proper quantity of reagent to add in order to precipitate the alanine. The weight of the salt is corrected for the small solubility to obtain the yield of alanine.

"To determine proline, advantage is taken of the insolubility of the proline compound of tetrathiocyanato-dianilidochromiato acid  $[\text{Cr}(\text{CNS})_4 \cdot (\text{C}_6\text{H}_5 \cdot \text{NH}_2)_2] \cdot \text{H}$  (rhodanilic acid). Again a solubility correction must be ascertained and applied.

"From the filtrate from the proline compound, hydroxyproline may be precipitated by the addition of Reinecke salt,  $[\text{Cr}(\text{CNS})_4(\text{NH}_3)_2] \cdot \text{NH}_4$ , together with pyridine. Purification of the hydroxyproline is difficult and the method is probably considerably less satisfactory as an analytical procedure than that employed for the other amino acids. Nevertheless higher yields, as compared with the results of other methods, were obtained.

"A consideration of the difficulties and uncertainties involved in these precipitation methods—particularly with respect to the incompleteness of precipitation and consequent necessity for the use of solubility corrections—has led during the past two years to the development of methods of a new type in Bergmann's laboratory<sup>121, 122, 123</sup>. If to a solution of an amino acid a reagent is added in quantity less than sufficient to precipitate all the amino acid, a precipitate will form that contains, when equilibrium is reached and on the simplest assumption of binary salt formation, equimolecular quantities of the reagent and the amino acid. The mother liquor will contain a certain concentration of the ions of both reagent and amino acid. If a known amount of reagent has been employed, the concentration of the reagent left in the mother liquor can be calculated and an equation set up for the solubility product of the two ions. If a second experiment with a different amount of reagent is then carried out, and the solubility product calculated, these two results can be equated. In the resulting expression, the only unknown is the molar concentration of the amino acid in the solution at the beginning. This calculation assumes that molar concentrations can be substituted for ionic concentrations, that activity coefficients do not change during the experiment, and, for success, requires rigid control of the physical conditions under which the experiment is conducted. The great advantage is that no knowledge of the solubility of the amino acid compound is required and, accordingly, the insidious errors are avoided that may creep in when the solubility of a compound in water is used as a correction for the solubility of the same compound in a

of the assumptions involved, quantitative recoveries of amino acids, either in pure solution or in mixtures, can be readily secured.

"New data have already been obtained in several cases by this method. For example, the glycine content of gelatin, previously given as 25.5 per cent, is now found to range from 26 to 27 per cent according to the source of the protein preparation. The proline, formerly 19.7 per cent, is now 17.5 per cent with a precision of about 2.5 per cent of this value. The revision downwards by the new method furnishes an illustration of the effect of errors in the earlier method.

"In order conveniently to apply this method, Bergmann and Stein<sup>124</sup> have proposed an entirely new reagent for certain amino acids, namely naphthalene- $\beta$ -sulfonic acid, the salts of which are referred to as nasylates, to distinguish them from the naphthylenesulfonyl derivatives. The selection of this substance, which yields salts that are only moderately insoluble, is possible because a definite and appreciable solubility is a distinct advantage in the application of the new solubility product method and simple binary salt formation with the amino acid is desirable. With this reagent, the determination of leucine, arginine, and probably other amino acids, is possible. The salts of isoleucine and valine are more soluble than those of leucine and, accordingly, these amino acids do not interfere with leucine determinations in proteins. The determinations of phenylalanine with this new reagent or some substance of analogous properties also appears to be a definite possibility.

"As yet this method has not been applied to more than a few proteins. But the accuracy with which solutions of amino acids, or simple mixtures of amino acids can be analyzed suggests that most valuable results are to be anticipated" (reference 14, p. 105-107).

Another method, recently introduced<sup>125, 126</sup>, which promises to be extremely powerful depends on the availability in concentrated form of isotopes such as deuterium and the nitrogen isotope of atomic weight 15. A pure sample of a given amino acid containing a known amount of deuterium or N<sup>15</sup>, in a stable position, is added to a completely hydrolyzed sample of a given protein. A sample of the same amino acid is then isolated in pure form from the hydrolysis mixture including the added amino acid. If the amount of added amino acid is  $x$  and its isotope content is  $C_0$  while the isotope concentration of the specimen of pure amino acid as finally isolated is  $C$ ; then the amount of the amino acid present in the original mixture— $y$ —is:

$$y = \frac{x(C_0 - 1)}{C - C_0}$$

(All isotope concentrations must be stated as excess above the amount normally present.) The great advantage of this method is that quantitative isolation of the amino acid is unnecessary. It is, on the other hand, extremely important that the sample of amino acid as finally isolated should be as pure as possible. This method has not yet been extensively employed but its possibilities appear to be very great. It should be noted, however, that while this method should permit very accurate determinations of the amino acid content of protein hydrolysates, it does not take account of any destruction of the amino acid which may have occurred during the process of hydrolysis. This limitation, of course, holds also for all other methods of analysis hitherto developed.

### Amino Acid Composition of Proteins

While awaiting the further development of these newer and more accurate methods we may none the less consider the present evidence of the amino acid composition of proteins. The proteins that have been repeatedly studied by a large number of investigators are not necessarily those that we now believe to be the best characterized chemical individuals. The organic chemists concerned with improvement in the methods for the isolation or estimation of the amino acid contents of proteins have often been satisfied with proteins readily prepared in large amount. Many of these have not been crystallized or characterized by the newer physical chemical methods. Indeed there is now evidence that many of them can be separated into more than one molecular species. Thus the protein fraction previously known as lactalbumin is now known to consist in part of a well defined crystalline globulin, lactoglobulin. Only fragmentary knowledge of its composition is as yet available. The analytical results on lactalbumin are therefore not considered here though they bore an interestingly close relation to those of casein<sup>90</sup>, Table 21. The results on casein are, however, included in Table 5, as are those of gliadin and zein, edestin and the coconut globulin, egg albumin and serum albumin, fibroin and the spider fibroin investigated by Emil Fischer. Although the results for egg and serum albumin are included because so many analytical measurements have been reported, it is to be expected that far better preparations of both of these crystalline proteins will presently be available and that the analytical results for them, as for other proteins when they have been further purified, will differ appreciably from those that are thus far available.

The results of analyses that have been reported are thus tabulated without any intention of considering many of them as yielding more than

tion of others, such as the dicarboxylic acids, suggest that these should be considered as minimal estimates. Higher yields will subsequently be reported, as has already occurred for the aspartic and glutamic acid contents of the proteins that are being studied by Chibnall.\* On the other hand, as Vickery has pointed out, certain of the results in the literature may prove to be somewhat too large because of the occlusion of impurities during precipitation<sup>14</sup>.

The sum of the percentile composition for several of the proteins for which results are tabulated exceeds 100 per cent, since the weight of the amino acid analyzed includes the water combined by the amino acid residue during hydrolysis. These results are recalculated (in Table 6) in terms of the weight of the amino acid residues for the few proteins whose analyses are most complete, so that if the amino acid composition of the protein were completely known the residue weights would equal 100 per cent.

Even in the cases for which this figure is approached, the tentative nature of these results should be borne in mind, for the analyses reported from different laboratories have often been carried out on quite different preparations. Moreover, the protein preparations employed and analyzed have rarely been shown to represent chemical individuals by any of the physical chemical methods upon which we may now rely for the purification and characterization of protein molecules.

It is none the less of interest to note the differences in the free groups of different proteins and to point out the importance that must be attached to the improvement of analytical procedures upon which our notions of the fine structure of the proteins must ultimately rest. The analyses on the basis of which their percentile compositions have been determined have been discussed elsewhere<sup>56, 90, 127, 128</sup>. Although it must be remembered that in many cases the number of groups of any kind may at this time only be estimates, since the analytical evidence regarding many of these residues is by no means satisfactory, there is none the less an advantage in considering the differences in protein molecules which emerge from the analyses at present available.

Zein and gliadin are classified as prolamines. That is to say, they are insoluble in water, but soluble in alcohol-water mixtures. Their compositions are known to over 85 per cent, and so it would appear to be significant that they contain very few basic amino acids and very few free carboxyl groups, consisting predominantly of glutamine and asparagine, of proline and of monoamino-monocarboxylic acids with paraffin



side chains. Even certain differences between zein and gliadin may be worth noting. Thus the former is poorer in amide and pyrrolidine groups, whereas the latter is poorer in the number of non-polar  $\text{CH}_2$  groups in paraffin side chains.

As compared with the small proportions of basic and carboxylic groups of zein, and the only slightly greater proportion of basic and free carboxylic groups of gliadin, egg albumin has far more basic and free carboxylic groups and is extremely soluble in water, being therefore classified as an albumin.

TABLE 6. WEIGHT PER CENT OF AMINO ACID RESIDUES IN PROTEINS

Amino acid	Molecular weight of amino acid residue	Weight per cent of amino acid residues				
		Insulin	Egg albumin	Edestin	Gelatin	Zein
Glycine.....	57.03		0	2.89	19.38	
Alanine.....	71.04		1.77	2.87	6.94	7.81
Serine.....	87.08	2.96		0.27	2.73	0.85
Threonine.....	101.10	2.26			1.19	
Valine.....	99.08		2.12	4.74		1.59
Leucine and isoleucine.....	113.08	25.88	9.24	18.03	6.12	21.56
Proline.....	97.08		3.50	3.46	16.62	7.63
Oxyproline.....	113.06			1.73	12.42	0.69
Phenylalanine.....	147.08		4.52	2.75	1.25	6.77
Methionine.....	131.18		4.60	2.10		2.07
Cystine.....	222.18	11.56	1.65	1.26	0.17	0.84
Tryptophane.....	186.08		1.21	1.33		
Tyrosine.....	163.08	11.25	3.58	4.09		5.31
Histidine.....	137.08	9.46	1.31	2.13	2.60	0.72
Arginine.....	156.08	2.73	5.07	15.03	7.78	1.43
Lysine.....	128.08	1.10	4.36	2.08	5.19	
Aspartic acid*.....	115.08		7.00	10.38	2.94	1.56
Glutamic acid*.....	129.08	13.80	4.79	1.83	1.70	
Hydroxyglutamic acid.....	145.08		1.21			2.22
Glutamine*.....	128.06	12.43	9.27	16.33	3.44	27.69
Per cent protein as amino acid residue: <i>W</i> .		93.43	65.20	93.30	90.47	88.74

The values in this table correspond to the values recorded in Table 5.

\* The assignment of the amide nitrogen to glutamine or asparagine is arbitrary in this calculation. For convenience all was assigned to glutamine whenever possible. The final result is independent of this choice.

The above proteins have been prepared in a variety of ways and studied by a variety of methods by different investigators. The criterion that a reproducible product was analyzed has therefore merely been the consistency of the analytical results. In contrast we may consider certain proteins which have more recently been crystallized and the analytical results on which they have for the most part been carried out by only a few investigators. Whereas these proteins are thus well characterized as chemi-

Physiological behavior may often depend upon a very small fragment of the molecule, the groups of which are arranged in a special configuration. In the case of hemoglobin and of cytochrome, of the hemocyanins, of the yellow enzymes, and of certain other proteins the reactive group may be considered as prosthetic, and is chemically quite unlike the amino acids. Even in these substances, however, the reactivity of the prosthetic group is fundamentally dependent on the nature of the amino acid residues in the protein molecule to which it is attached. In the case of insulin, however, as we have seen, investigation has thus far revealed no components of the molecule other than the normally occurring amino acids, and the specificity of physiological behavior of this molecule<sup>56, 130, 131</sup> would thus appear to depend, as in the case of many enzymes, in large part upon the arrangement in space of charged and uncharged, polar and non-polar amino acid side chains.<sup>132</sup>

described a method of determining hydroxyl groups in proteins. The results of the analysis of the protein from which Van Slyke and his associates prepared the hydroxyl group reagent,  $\alpha$ -hydroxy- $\beta$ -methylvaleryl-L-proline, are shown in Table I. The results of the analysis of the protein from which Van Slyke and his associates prepared the hydroxyl group reagent,  $\alpha$ -hydroxy- $\beta$ -methylvaleryl-L-proline, are shown in Table I. The results of the analysis of the protein from which Van Slyke and his associates prepared the hydroxyl group reagent,  $\alpha$ -hydroxy- $\beta$ -methylvaleryl-L-proline, are shown in Table I.

## Chapter 16

### Density and Apparent Specific Volume of Proteins

BY EDWIN J. COHN AND JOHN T. EDSALL

The apparent molal volumes of amino acids and related compounds have been considered in detail in Chapter 7. From the discussion given there, it is apparent that particular atomic groupings—such as the  $\text{CH}_2$ , the  $\text{CONH}$ , or the  $\text{OH}$  group—contribute characteristic increments to the apparent molal volume of a molecule, these increments being nearly independent of the nature of the other groups present. Similarly, the apparent molal volume of a simple protein may be computed from the characteristic volume increments of the amino acid residues of which it is composed, if the composition of the protein is known.<sup>1,2</sup> Denote the glycyl residue as component 1, alanyl as 2, etc. Then, if the protein contains  $n_i$  moles of the  $i$ 'th amino acid residue (of molecular weight  $M_i$ , and apparent molal volume  $\Phi_i$ , the apparent molal volume of the protein,  $\Phi_p$ , is:

$$\Phi_p = \sum_i n_i \Phi_i \quad (1)$$

This equation neglects the "covolume" of Traube (see Chapter 7), but the magnitude of the covolume is only about 14 cc. This is negligible in comparison with the molal volume of a protein molecule, which is always many thousands of cc. Equation 1 also neglects the correction for electrostriction due to the charged groups of the protein; this correction is small, and is further considered below.

The molecular weight of the protein,  $M_p$ , is

$$M_p = \sum_i n_i M_i \quad (2)$$

Hence, subject to the qualifications stated above, the apparent specific volume of the protein is

$$V_p = \frac{1}{\rho_p} \quad \frac{\Phi_p}{M_p} = \frac{\sum n_i \Phi_i}{\sum n_i M_i} \quad (3)$$

$\rho_p$ , the reciprocal of the apparent specific volume, may be called the "apparent density" of the protein in solution.

in determining molecular weights of proteins by sedimentation and diffusion (Chapter 19), since the sedimentation velocity of a protein is proportional to the difference in density between the protein and the surrounding medium. The discussion given above indicates that the density of a protein should be determined by the relative numbers, and the densities, of the different amino acid residues of which it is composed. Thus the densities (or specific volumes) of proteins furnish an index of their composition.

Equation 3 involves the molecular weight of the protein, but it is not necessary to know this in order to evaluate the density from the composition. If  $W_i$  is the weight per cent of the  $i$ 'th amino acid residue in the protein, and  $V_i$  the specific volume of this residue ( $V_i = \Phi_i/M_i$ ) then the volume of 100 gm of protein is  $\Sigma V_i W_i$  and the apparent specific volume is

$$V_p = \frac{\Sigma V_i W_i}{100} \quad (4)^*$$

Knowledge of the  $W_i$ 's depends on the amino acid analyses discussed in Chapter 15. For no protein is the amino acid analysis complete, so that some of the  $W_i$ 's are still unknown. If we take account only of the analytically determined residues, we may calculate the apparent specific volume due to these known residues by the equation

$$V_p = \frac{\Sigma V_i W_i}{\Sigma W_i} \quad (4a)$$

the summation extending over all the analytically determined  $W_i$ 's. If all the  $W_i$ 's are known, 4a becomes identical with 4, since then  $\Sigma W_i = 100$ . For several proteins the amino acid analysis is so nearly complete that the specific volume may be fixed within narrow limits by the analytical data, provided the specific volumes of the residues are known. We shall now consider the evaluation of these volumes.

#### The Specific Volumes of Amino Acid Residues

**The Glycyl Residue,  $-\text{NH}\cdot\text{CH}_2\cdot\text{CO}-$ .** Two different methods of evaluating the molal volume of this residue may be employed. On the one hand, the residue of glycine may be considered as the sum of the  $\text{CH}_2$  and the  $\text{CO}\cdot\text{NH}$  groups. Taking 16.3 cc as the volume of the  $\text{CH}_2$  group and 20 cc as the volume of the  $\text{CONH}$  group, the molal volume of the glycine residue is calculated as 36.3 cc. The alanine residue would on this basis be greater by 16.3 cc; the valine residue by three times 16.3, and the leucine residue by four times 16.3 cc.

the apparent molal volumes of glycine peptides, provided the electrostriction due to the respective molecules is the same. This would appear to

TABLE 1. MOLAL VOLUMES OF AMINO ACIDS AND THE CALCULATION OF THE SPECIFIC VOLUMES OF AMINO ACID RESIDUES

	Molal volume of amino acid observed cc.	Molal volume of residue (Observed volume -7.4)* cc.	(Calculated from groups)† cc.	Specific volume of residue
Glycine.....	43.5	36.1	36.3	0.64
Alanine.....	60.6	53.2	52.6	.74
Serine.....	60.8	53.4	54.9	.63
Threonine.....			71.2	.70
$\alpha$ -Aminobutyric acid.....	76.5	69.1	68.9	.81
$\alpha$ -Aminovaleric acid.....	92.7	85.3	85.2	.86
$\alpha$ -Aminocaproic acid.....	108.4	101.0	101.5	.90
Proline.....	81.0	73.6		.76
Oxyproline.....	84.4	77.0		.68
Phenylalanine.....	121.3	113.9		.77
Methionine.....	105.1	97.7		.75
Cystine.....			135.4	.61
Tryptophane.....	144.1	136.7		.74
Tyrosine ‡.....	123.6	116.2		.71
Histidine†.....	99.3	91.9		.67
Arginine§.....			109.1	.70
Lysine§.....	108.5	105.8	105.1	.82
Aspartic acid.....	74.1	66.7	68.4	.60
Glutamic acid.....			84.7	.66
Oxyglutamic acid.....			87.0	.60
Asparagine.....	78.0	70.6	72.6	.62
Glutamine.....	93.9	86.3	88.9	.67

The italicized values of the molal volumes of the residues have been employed in computing the specific volumes used in the last column.

Cohn, McMeekin, Edsall and Blanchard [*J. Am. Chem. Soc.*, 56, 784 (1934)] give slightly lower values for the molal volumes of isovaleric and isocaproic acids, but a slightly higher value for isobutyric acid, than for their normal isomers. Thus for  $\alpha$ -aminocaproic they give 108.4, for  $\alpha$ -aminoisocaproic 107.5, whereas Dalton and Schmidt [*J. Biol. Chem.*, 109, 241 (1935)] give 105.5 for isoleucine. They also give 106.4 for methionine, 128.7 for phenylalanine and 61.6 for serine.

\* For the derivation of the factor 7.4 cc. see the text, equations 5 and 6

†  $\text{NH}_2 = 7.7$ ;  $\text{CH}_2 = 16.3$ ;  $\text{COOH} = 18.9$ ;  $\text{CONH} = 20.0$ , etc.

‡ Calculated from phenylalanine.

§ These are  $\epsilon$ -amino acids. The electrostriction is therefore taken as 18 cc. per mole instead of 13.3.

TABLE 2. APPARENT MOLAL VOLUMES OF HYDANTOIC ACIDS

Solute	$\text{CH}_2$ groups (cc.)	$\text{CONH}$ groups (cc.)	$\text{NH}_2 +$ $\text{COOH}$ groups (cc.)	Apparent molal volume (cc.) Calcu- lated	Ob- served
Hydantoic acid.....	16.3	20.0	40.7	77.0	77.6
$\alpha$ -Alanine hydantoic acid.....	32.6	20.0	40.7	93.3	94.2
$\beta$ -Alanine hydantoic acid.....	32.6	20.0	40.7	93.3	95.8
$\alpha$ -Aminocaproic hydantoic acid.....	81.5	20.0	40.7	142.2	146.4
$\epsilon$ -Aminocaproic hydantoic acid.....	81.5	20.0	40.7	142.2	146.4
Diglycine hydantoic acid.....	32.6	40.0	40.7	113.3	112.4
Triglycine hydantoic acid.....	48.9	60.0	40.7	149.6	149.5

both are  $\alpha$ -amino acids (Tables 1 and 2).<sup>\*</sup> Even more satisfactory computations can be made by comparison of hydantoic acid with the hydantoic acids of glycine peptides or of glycolamide with glycyglycolamide. The results for the hydantoic acids, given in Table 2, confirm the estimate of the volume increment of the  $\text{CH}_2\text{CONH}$  group derived from the sum of the  $\text{CH}_2 + \text{CONH}$  groups. The latter estimate must be considered more accurate, however, since it is based on many more measurements.

**Correction for Covolume and Electrostriction.** The molal volumes of amino acid residues might also be calculated by subtracting a constant correction factor from the observed molal volumes of all  $\alpha$ -amino acids. This factor would take account of the water eliminated when amino acids are bound in peptide linkage, and also of the covolume (Chapter 7). When this factor is taken as 7.4 cc/mole for all  $\alpha$ -amino acids, the molal volumes listed in Table 1 for amino acid residues are obtained. The results for glycine, alanine and other  $\alpha$ -amino acids are in very satisfactory agreement with the values calculated from the volumes of the groups, which are also listed in Table 1. The value 7.4 might have been adopted empirically because of this agreement, but it may also be deduced as follows. The observed apparent molal volume ( $\Phi$ ) of an amino acid equals the sum of the volume increments due to the constituent groups, plus the "covolume" of Traube, minus the electrostriction of the solvent ( $E$ )<sup>†</sup>.

For  $\alpha$ -amino acids we may take  $E$  as 13.3 cc/mole, from the data given in Chapter 7. Thus, for glycine, we may estimate the covolume from the relation:

$$\begin{aligned}\text{Covolume} &= \Phi(\text{Glycine}) + E - V_{\text{NH}_2} - V_{\text{CH}_2} - V_{\text{COOH}} \\ &= 43.5 + 13.3 - 7.7 - 16.3 - 18.9 \\ &= 14.1 \text{ cc}\end{aligned}\tag{5}$$

When a glycine molecule is removed from solution, and incorporated as a glycy residue in a protein molecule, three volume changes occur: (1) a molecule of water is eliminated; this reduces the volume increments of the constituent groups (employing Traube's figures for H and O) by 6.6 cc; (2) the charges on the terminal groups are eliminated; hence the electrostriction vanishes; this gives rise to an expansion of 13.3 cc/mole; (3) the covolume term becomes negligible, since the covolume for the whole protein molecule is only 14 cc, and this is a negligible fraction of the total volume of the protein. Thus this term gives rise to a contraction of 14.1 cc (see Equation 5). Thus considering these three effects:

$$\Phi(\text{glycyl residue in protein}) - \Phi(\text{glycine in solution}) = -6.6 + 13.3 - 14.1$$

which is the factor used in obtaining the values in Table 2. The same reasoning applies to other  $\alpha$ -amino acids.

Certain minor discrepancies between Columns 3 and 4 may be noted. Thus the molal volumes of the asparagine and glutamine residues calculated in Column 3 are smaller by 2 cc/mole than those given in Column 4. The same is true of aspartic acid. It is possible that the value 13.3 adopted for electrostriction is too small. This is certainly true in the case of lysine. In general, isomers with different arrangements in the side chains give slightly different values of  $\Phi$ , and this will give rise to differences in the electrostriction values calculated from these data (Chapter 7). Furthermore it must be remembered that the apparent molal volume of either an amino acid or a protein varies with the concentration.

The molal volume of any residue divided by its equivalent weight yields its specific volume, and this is given in the final column of Table 1. (The value of the molal volume from which it is derived is italicized.) The specific volume of the COOH group is 0.420, that of the CONH group 0.465, that of the NH<sub>2</sub> group 0.481, and that of the CH<sub>2</sub> group 1.63. Despite the wide variation between these results the specific volumes of amino acid residues vary only from .60 for aspartic acid, .64 for glycine and .67 for histidine to .82 for lysine, .86 for norvaline and .90 for norleucine.

### Specific Volumes of Proteins

The specific volumes of all proteins must fall within the upper and lower limits for the amino acid residues. Those of proteins which yield large amounts of glycine and aspartic or glutamic acid on hydrolysis should be smallest, those containing large amounts of valine, leucine or lysine, largest. Even with the incomplete yields of amino acids thus far obtained estimates may be made of the specific volumes of proteins using Equation 4a and the data of Tables 1 and 2. This has been attempted in Table 3. The necessary analytical data are taken from Chapter 15, Table 6. The volume per cent  $V_i/W_i$  of each known amino acid residue in the protein is recorded in Table 3. Thus the volume of insulin represented by leucine residues is calculated to be more than 23 volumes per cent of the protein and of glutamine, nearly 10 volumes per cent.

The calculated specific volumes of insulin, egg albumin and edestin are all equal to 0.73 to 0.74. Gelatin, with a very high glycine content, is estimated to have a smaller specific volume, 0.71, and zein, with a large leucine content, a slightly larger specific volume.

These calculations must be considered tentative. The proteins investigated have not in every case been demonstrated to be chemical

estimates of the specific volumes of the acid residues. Such calculations should, however, enable us to determine how far the specific volumes of the proteins are determined by those of the groups of which proteins are constituted.

TABLE 3. VOLUME FRACTIONS OF AMINO ACID RESIDUES IN PROTEINS AND CALCULATION OF THEIR SPECIFIC VOLUMES

Amino acid	Specific volume of amino acid residue $V$	Volume per cent of amino acid residue: $VW$				
		Insulin	Egg albumin	Edestin	Gelatin	Zein
Glycine.....	0.64			1.85	12.40	
Alanine.....	.74		1.31	2.12	5.13	5.78
Serine.....	.63	1.86		0.17	1.72	0.53
Threonine.....	.70	1.58			0.83	
Valine.....	.86		1.82	4.08		1.37
Leucine and isoleucine.....	.90	23.29	8.32	16.23	5.51	19.40
Proline.....	.76		2.66	2.63	12.63	5.80
Oxyproline.....	.68			1.18	8.44	0.47
Phenylalanine.....	.77		3.48	2.12	0.96	5.21
Methionine.....	.75		3.45	1.57		1.55
Cystine.....	.61	7.05	1.01	0.77	0.10	0.51
Tryptophane.....	.74		0.90	0.98		
Tyrosine.....	.71	7.98	2.54	2.90		3.77
Histidine.....	.67	6.34	0.88	1.43	1.74	0.48
Arginine.....	.70	1.91	3.55	10.52	5.45	1.00
Lysine.....	.82	0.90	3.57	1.71	4.25	
Aspartic acid*.....	.60		4.20	6.23	1.73	0.93
Glutamic acid*.....	.66	9.10	3.16	1.21	1.12	
Hydroxyglutamic acid.....	.60		0.73			1.33
Glutamine*.....	.67	8.33	6.21	10.94	2.30	18.55
Volume per cent of amino acid residues..... $\Sigma V_i W_i$		68.34	47.79	68.64	64.31	66.68
Weight per cent protein of amino acid residues..... $\Sigma W_i$		93.43	65.20	93.30	90.47	88.74
Specific volume						
Calculated..... $\Sigma V_i W_i / \Sigma W_i$		0.73	0.73	0.74	0.71	0.75
Observed.....		0.749	0.749	0.744		

\* The assignment of the amide nitrogen to glutamine or asparagine is arbitrary in this calculation. For convenience all was assigned to glutamine whenever possible. The final result is independent of this choice.

### Partial Specific Volumes of Proteins in Solution

The density of proteins is far greater than that of water and their partial specific volumes are therefore far smaller. The partial specific volume is generally estimated pycnometrically in dilute aqueous or electrolyte



The difficulty of measuring satisfactorily the dry weight of protein samples has often been pointed out and depends upon their great affinity for small amounts of water. Generally small amounts, less than 200 mgs, are dried near 100° for 24 to 48 hours, and weighed in closed weighing bottles. Errors due to residual water are always possible. They could be estimated by freezing point lowering in organic solvents in which the protein is soluble<sup>3</sup>. Protein concentration is often estimated by nitrogen determination, but the nitrogen factor employed in the calculation is itself determined from a nitrogen analysis on the dried protein.

The partial specific volumes of a few amino acids and peptides, and of a few related polar molecules are given in Table 4. Urea, despite the absence of the non-polar CH<sub>2</sub> group, is by no means as dense as glycine or

TABLE 4

Substance	Apparent specific volume		$\Delta$
	In solid state	In aqueous solution	
Acetamide.....	0.863 (1)	0.931 (5)	+0.068
Propionamide.....	0.960 (1)	0.972 (5)	+0.012
Butyramide.....	0.969 (1)	0.999 (5)	+0.030
Acetylurea.....	0.704 (2)	0.723 (2)	+0.019
Urea.....	0.749 (1)	0.783 (6, 7)	-0.011
Methyl hydantoic acid.....	0.715 (3)	0.713 (3)	-0.002
Glycylglycine.....	0.644 (3)	0.584 (3)	-0.060
L-Asparagine.....	0.648 (1)	0.591 (3)	-0.057
Glycine.....	0.622 (4)	0.583 (8, 9)	-0.039
D-Alanine.....	0.714 (4)	0.684 (8, 9)	-0.030
D-Valine.....	0.813 (4)	0.770 (9)	-0.043
L-Leucine.....	0.858 (4)	0.817 (8, 9)	-0.041

From Cohn, E. J., *Ann. Rev. Biochem.*, **4**, 93 (1935).

- (1) *Physical Properties of Organic Compounds*, Vol. 1.
- (2) *Physical Properties of Organic Compounds*, Vol. 1.
- (3) *Physical Properties of Organic Compounds*, Vol. 1.
- (4) *Physical Properties of Organic Compounds*, Vol. 1.
- (5) *Physical Properties of Organic Compounds*, Vol. 1.
- (6) *Physical Properties of Organic Compounds*, Vol. 1.
- (7) *Physical Properties of Organic Compounds*, Vol. 1.
- (8) Dalton, J. B., and Schmidt, C. L. A., *J. Biol. Chem.*, **103**, 549 (1933).
- (9) Cohn, E. J., McMeekin, T. L., Edsall, J. T., and Blanchard, M. H., *J. Am. Chem. Soc.*, **56**, 784 (1934).

glycylglycine. Methyl hydantoic acid is an isomer of asparagine and of glycylglycine, but has a much larger apparent specific volume, both in the solid state and in solution. The much greater density of the two latter isomers thus reflects their dipolar ionic nature. This is true also of the specific volumes of the  $\alpha$ -amino acids. That of glycine and its peptides is far smaller and that of leucine far larger than the values characteristic of proteins. The change in apparent specific volume on solution ( $\Delta$ ) for these various molecules is also given in the last column of Table 4. For the polar molecules which are not dipolar ions  $\Delta$  has either a positive or a

of these molecules. (Chapter 7). The very considerable negative values of  $\Delta$  for the amino acids should be compared with the values given for proteins in Table 7 for two proteins in which the change of specific volume on solution is zero within the limits of experimental error.

The experimentally determined partial specific volumes of the proteins are all nearly the same and in the neighborhood of 0.74 or 0.75.\* This result is essentially what should be expected even in the present incomplete state of our knowledge of protein composition.

### Density of Protein Crystals

The most important studies on protein crystals are those of Adair and Adair<sup>4</sup>. They suspended the crystals in solvents of approximately equal

TABLE 5. DENSITIES OF CRYSTALS OF EDESTIN, HORSE SERUM ALBUMIN, EGG ALBUMIN, AND OF HORSE SERUM EUGLOBULIN  
 $D$  = density reduced to 20°.  $T$  = temperature of measurement.  $W$  = provisional estimate of hydration in gm water per gm dry protein

Protein	Medium	Temperature (°C.)	$D$	$W$
Edestin	Saturated ammonium sulfate and sucrose	21.1	1.317	0.063
Edestin	Phosphate pH 5.0	22.0	1.288	0.143
Edestin	Citrate pH 8.3	24.0	1.288	0.143
Edestin	Citrate pH 6.6	20.9	1.290	0.137
Edestin	Citrate pH 4.3	24.4	1.308	0.086
Serum albumin	Half-saturated $(\text{NH}_4)_2\text{SO}_4$ and sucrose pH 4.8	18.0	1.276	0.192
Serum albumin	Two-thirds saturated $(\text{NH}_4)_2\text{SO}_4$ and sucrose pH 4.8	18.0	1.279	0.182
Serum albumin	Citrate pH 6.6	20.8	1.246	0.304
Serum albumin	Phosphate pH 5.0	20.3	1.237	0.344
Serum albumin	Phosphate pH 3.2	21.5	1.278	0.185
Egg albumin	Phosphate pH 5.0	19.0	1.239	0.324
Egg albumin	Phosphate pH 3.2	20.0	1.268	0.208
Globulin	Phosphate pH 5.0.....	20.0	1.236	0.331

From G. S. Adair and M. E. Adair, *Proc. Roy. Soc. London*, **B120**, 430 (1936).

density, and centrifuged the suspension, observing whether the crystals rose or sank. The solvents used were concentrated solutions of ammonium sulfate, of phosphate or citrate buffers, and of sucrose, or mixtures of these. By dilution with water, the density of the solvents could be varied through a suitable range; some being denser, some less dense, than the protein crystals. The density of the crystals was then determined by interpolation from the results of a series of centrifuge experiments.

Some of their density values are listed in Tables 5 and 6†.

It is immediately apparent from these results that the density of protein crystals is always less than the "apparent density" of proteins in solution,

among the protein crystals studied, has a density approaching this value. All of the crystals of animal proteins are less dense than edestin, and the density of the crystals varies markedly with the nature of the suspension medium. Adair and Adair showed that these density variations were reversible; a sample of protein crystals could be transferred from medium A to medium B, in which the crystal density was found to differ from that in A; when the crystals were brought back into medium A, their density rapidly assumed the same value it had had originally in the same medium.

Superficially there is an analogy here to the behavior of the amino acids on solution in water; for the density of an amino acid crystal is generally less than the reciprocal of the apparent specific volume of the amino acid in water (see Chapters 7 and 8). This is not generally true of organic compounds, and the distinctive behavior of the amino acids may readily

TABLE 6. DENSITIES OF CRYSTALS OF HORSE AND OF SHEEP HEMOGLOBIN

Species	Medium	Temperature (°C.)	<i>D</i>	<i>W</i>
Horse	Half-saturated $(\text{NH}_4)_2\text{SO}_4$ + sucrose	19.0	1.269	0.166
Horse	$(\text{NH}_4)_2\text{SO}_4$ + sucrose pH 6.8	17.9	1.264	0.183
Sheep	$(\text{NH}_4)_2\text{SO}_4$ + sucrose pH 6.8	20.8	1.264	0.183
Horse	0.01 M phosphate + sucrose	17.9	1.267	0.172
Horse	$\text{NaH}_2\text{PO}_4$ + sucrose	17.9	1.252	0.227
Horse	Phosphate pH 5.0	20.0	1.225	0.344
Sheep	Phosphate pH 5.0	19.7	1.226	0.338
Sheep	Citrate pH 6.6	20.4	1.231	0.315
Sheep	Saturated $(\text{NH}_4)_2\text{SO}_4$ + $\text{H}_2\text{O}$	22.0	1.242	0.267
Horse	Saturated $\text{MgSO}_4$ + $\text{H}_2\text{O}$	24.5	1.238	0.284

From G. S. Adair and M. E. Adair, *Proc. Roy. Soc. London*, B120, 431 (1936).

be explained by the electrostriction of the solvent due to their charged groups. A similar explanation of the difference between protein crystals and proteins in solution, however, is inadequate to explain the facts. We may calculate the electrostriction, for instance, in egg albumin. From the recent data of Cannan, Kibrick and Palmer<sup>5</sup>, egg albumin at the isoelectric point contains 91 positively charged groups per 100 kg of dry egg albumin, and an equal number of negative charges. The apparent specific volume is 0.75, hence the apparent volume per 100 kg is 75,000 cc. The electrostriction due to one pair of charged groups, when the charges are widely separated, is 20 cc/mole. Hence 91 pairs give  $91 \times 20 = 1820$  cc electrostriction per 100 kg of egg albumin, which is 2.4% of the experimentally determined apparent volume. The density of egg albumin crystals (Table 5), however, was found to be as low as 1.24, or about 7% less than the apparent density in solution; so that electrostriction is not an adequate explanation of the observed facts.

proteins by weighing them in benzene. They obtained densities between 1.269 for crystalline egg albumin and 1.318 for casein; the values were always lower than the apparent density in solution. However, Adair and Adair showed that edestin and horse hemoglobin had identical densities in suspension and solution, within the experimental error (Table 7). The contraction on solution, which would be expected if there were a large electrostriction effect, is not present. It seems possible that, in the work of Chick and Martin, the densities obtained were too low, because of small occluded air bubbles, or because their crystalline egg albumin may not have been wholly dry.

An additional factor, probably much more important than electrostriction, is hydration of the crystals. Sørensen<sup>7</sup> in 1917 studied suspensions of egg albumin crystals surrounded by ammonium sulfate solution, and the filtrates of the mother liquor from such suspensions. Analysis

TABLE 7. PARTIAL SPECIFIC VOLUME OF DRY PROTEIN IN SUSPENSION AND IN SOLUTION

Protein .....	Horse hemoglobin	Edestin
Temperature (°C) .....	1.0	21.9
Apparent density of dry protein in the form of solid suspension in distilled water, wholly crystalline in the case of hemoglobin .....	1.344	1.349
Partial specific volume of dry protein in the form of suspension ..	0.744	0.741
Partial specific volume of dry protein in solution .....	0.742	0.745
Contraction in ml/gm crystalline protein dissolved .....	0.002	-0.003

From G. S. Adair and M. E. Adair, *Proc. Roy. Soc. London*, B120, 441 (1935).

of the crystal suspension, and of the filtrate solution—assuming the latter to have the same composition as the mother liquor surrounding the crystals—permitted the calculation of the water held in the crystals. If  $r_a$  denotes grams of anhydrous protein per gram protein nitrogen and  $r$  denotes gram hydrated protein per gram protein nitrogen, then

$$W = \frac{r - r_a}{r_a} \quad (7)$$

where  $W$  is apparent hydration in gram water per gram anhydrous protein. The method of calculating  $r$  was described by Sørensen in detail. He gave  $W$  for egg albumin as 0.22 gm.

Adair and Adair adopted a different method for the calculation of  $W$ , which depends on the relative densities of protein crystal, anhydrous protein, and water. Their analysis of the problem leads to the formula

$$(D_c - D) / D,$$

where  $D$  = density of crystal,  $D_p$  = density of anhydrous protein (reciprocal of apparent specific volume in solution), and  $D_1$  = density of water. The values of  $W$  so calculated are shown in Tables 5 and 6. They range from 0.063, for edestin in saturated ammonium sulfate and sucrose, to 0.344 for serum albumin in phosphate buffer at pH 5 and for horse hemoglobin in the same solvent. There are some uncertain elements in the calculation; particularly (as Adair clearly pointed out) we cannot determine the exact proportions in which water and the other components of the solvent medium are incorporated in the protein crystal. Sørensen's work showed clearly that protein crystals, in equilibrium with ammonium sulfate solution, contain relatively far more water and less ammonium sulfate than the mother liquor; but the exact proportions of the two cannot as yet be specified.

✓ Since this chapter was written, McMeekin and Warner<sup>8</sup> have determined directly the hydration of crystals of  $\beta$ -lactoglobulin. This protein can be crystallized either from water or from concentrated salt solutions, and crystals 2 or 3 mm long, and 1 mm in each of the other dimensions, can readily be obtained. McMeekin and Warner wiped adhering mother liquor off the crystal by placing it between two smooth pieces of cotton flannel, set the crystal on the pan of a sensitive balance, and determined the loss of weight as a function of time, on drying at room temperature in air. The weight of the crystal at zero time was determined by extrapolation. The final dry weight of the crystal was obtained by drying *in vacuo* over phosphorus pentoxide, or in a vacuum oven at 80°. Control experiments with crystals of  $(\text{NH}_4)_2\text{SO}_4$ ,  $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$ , and  $\text{Na}_2\text{SO}_4 \cdot 10\text{H}_2\text{O}$  indicated that the method gives accurate values for weight at zero time ✓ and for hydration.

Different experiments gave very consistent results. The loss of water was on the average 0.66 gram, per gram of air dried protein; and 0.84 gram per gram of oven dried protein. These are of the same order as the value of 0.54 gram of water per gram of air dried protein calculated from the data of Crowfoot<sup>9</sup>, and much higher than the values calculated by the method of Sørensen or that of Adair.

The amount of water associated with one gram of protein remained unchanged when the salt free crystals were placed in concentrated ammonium sulfate solutions. Ammonium sulfate, however, was found to diffuse into the crystal. The ratio of ammonium sulfate to water in the crystal was found by direct analysis to be on the average 82% of that in the mother liquor, for mother liquors containing 28.9 to 31.6% of ammonium sulfate by

The density of wet salt-free crystals was found to be 1.146, by immersion in bromobenzene-xylene mixtures of a series of different densities. After the crystals had remained 24 hours in the immersion medium, their density had risen to that of dry crystals (1.260). On the other hand, the density of wet crystals equilibrated with a solution of saturated ammonium sulfate with added sodium sulfate was 1.240, showing the great increase in density of wet crystals due to the uptake of salt. The hydration values for wet crystals calculated from these density values agreed very well with those directly determined.

These experiments furnish the most direct evidence yet obtained concerning hydration. They raise serious questions concerning the validity of the assumptions made by Sørensen and by Adair in their indirect calculations of protein hydration; if these assumptions are used, employing McMeekin's data for the density of  $\beta$ -lactoglobulin crystals in equilibrium with salt solutions, the values of hydration thus calculated are less than half as great as those directly determined by the drying experiments. As yet no protein except  $\beta$ -lactoglobulin has been studied by this direct method, but it seems probable that the hydration figures for other protein crystals determined by this method may prove to be higher than those previously inferred by calculation.

As we shall see later in Chapters 18, 19, 21 and 22, the hydration of protein crystals in solution is a factor of great importance in calculations of their size and shape. The hydration of a protein molecule in solution is, however, an essentially different thing from the hydration of a crystal. We may define it as the number of grams of water carried by one gram of protein when it moves through the solvent, as in sedimentation, diffusion or electrophoresis experiments. Data on the hydration of protein crystals give only indirect information concerning hydration in this sense. Knowledge of hydration in solution is still very meager, and remains an uncertain point in the interpretation of all methods for estimating the shape of protein molecules. The data on protein crystals still remain a valuable guide to the approximate nature of the effects to be expected in solutions.

## Chapter 17

### Osmotic Pressure and Molecular Weight of Proteins

BY JOHN T. EDSALL AND EDWIN J. COHN

The measurement of the colligative properties of solutions—freezing point depression, boiling point elevation, and osmotic pressure—has long been a standard procedure in the determination of molecular weights. All these measurements are based on the fact that the activity of the solvent is decreased by the addition of small quantities of solute, and that in very dilute solutions the activity of the solvent is proportional to its mole fraction. In the study of osmotic pressure it is necessary that the solution to be studied be in equilibrium with the pure solvent, and that the two solutions be separated by a membrane permeable to the solvent and impermeable to the solute<sup>1</sup>. In the absence of a pressure difference between the solutions in the inside and outside of the membrane, it is clear that equilibrium cannot exist, for the activity of the pure solvent at a given pressure and temperature must be greater than the activity of the same solvent when it is diluted by the presence of solute molecules. Equilibrium, however, can be attained through the application of a pressure difference between the solutions inside and outside the membrane. It is possible to obtain equilibrium by the application of positive hydrostatic pressure to the solution within the membrane, or of a negative pressure to the solution outside. Both methods have been used in practice. In either case the attainment of equilibrium depends upon the fact that the vapor pressure of any constituent of any phase increases with hydrostatic pressure according to the thermodynamic relation.

$$\frac{\partial \mu_k}{\partial P} = RT \frac{\partial \ln a_k}{\partial P} = \bar{V}_k \quad (1)$$

Here  $P$  is the hydrostatic pressure and  $\mu_k$  is the chemical potential of the  $k$ 'th component,  $a_k$  is its activity, and  $\bar{V}_k$  its partial molal volume. Thus the system is in equilibrium if the increase in the activity of the solvent, produced by the hydrostatic pressure, exactly balances the decrease in

have already been set forth.\* They lead, in the case of a sufficiently dilute solution, to the well known limiting law of van't Hoff

$$\lim_{c \rightarrow 0} \left( \frac{\pi}{C} \right) = RT \quad (2)$$

Here  $\pi = P'' - P'$  is the osmotic pressure, that is, the difference in hydrostatic pressure between the inner and outer solutions at equilibrium, and  $C$  is the concentration of the solute in moles per liter of solution.

In the past much emphasis has been laid upon the formal analogy between this equation and the perfect gas law. The osmotic pressure in a very dilute solution is indeed the same as the pressure which would be exerted at the same temperature by gas molecules at the same volume concentration as the solute molecules. There is, however, a profound difference in the mechanisms underlying the two phenomena. Gas pressure is due to the impact of the gas molecules on the walls of the container. Osmotic pressure is due to the decrease of the activity of the solvent due to the presence of the solute, and to the counterbalancing of this decrease by the applied pressure. The magnitude of this pressure is determined by the difference in mole fraction, or more generally in the activity, at constant pressure, of the solvent between the two phases. It has no direct relation to the impact of the solute molecules upon the walls of the membrane in which they are contained. From the underlying thermodynamic relations given in Chapter 3, indeed, it would be legitimate to write van't Hoff's equation in the alternative form

$$\lim_{m \rightarrow 0} \left( \frac{\pi}{m} \right) = RT \quad (3)$$

where  $m$  is the concentration of the solute in moles per liter of solvent. At infinite dilution, of course,  $C$  and  $m$  are identical. It is generally found in practice that equation 3 fits the data on protein solutions, and on most other solutions, better than does equation 2.

Thus, if  $g$  is the concentration in grams per liter of the solute to which the membrane is impermeable, and  $M$  is the molecular weight of the solute, then

$$M = RT \lim_{g \rightarrow 0} \left( \frac{g}{\pi} \right) \quad (4)$$

This equation is true only as a limiting relation valid at low concentrations. At higher concentrations it is found that the osmotic pressure almost always rises more rapidly than would be predicted by equations 2 or 3.



weights of proteins is far superior to the method of freezing point depressions or boiling point elevation. The last, indeed, would be totally inapplicable to most proteins, since they are denatured at the boiling point. The freezing point method is scarcely applicable for accurate determinations<sup>2</sup>, since small amounts of impurities of low molecular weight would seriously falsify the results. Furthermore, even if proteins could be obtained in pure condition, their freezing point depression in dilute aqueous solutions would be so small that the results would be inaccurate. Thus a 1% solution of egg albumin which is approximately  $2.4 \times 10^{-4}$  molar would give a freezing point depression, in water, of less than  $0.0005^\circ$ . On the other hand, such a solution gives an osmotic pressure in the neighborhood of 5 mm mercury or 65–70 mm of water, a quantity which is susceptible to reasonably accurate measurements. Furthermore, since impurities of low molecular weight can readily pass the membrane, the presence of small amounts of such impurities has no appreciable effect on the osmotic pressure due to the protein. It is in practice not very difficult to prepare membranes which are impermeable to proteins, but which are permeable to all smaller molecules in solution. It is worthy of remark that the study of osmotic pressure has arisen largely from the work of biologists. This is hardly surprising, in view of the great importance of semipermeable membranes in biology. It was indeed largely the work of the botanists Pfeffer<sup>3</sup> and de Vries<sup>4</sup>, on the relation of osmotic pressure to the concentrations of solutes, which enabled van't Hoff to formulate his celebrated law, which is embodied in equation 2.

<sup>4</sup> In principle the determination of molecular weights from osmotic pressure measurements by equation 4 is very simple, but in practice many difficulties arise. (1) The osmotic pressures of protein solutions are fairly small, particularly for large proteins. In very dilute solutions, for which alone van't Hoff's law is valid, it is difficult to obtain accurate readings. In concentrated solutions, the ratio of osmotic pressure to molar concentration is usually larger—often much larger—than that given by van't Hoff's law. (2) The earlier workers in this field, notably Sørensen<sup>5</sup> and Adair<sup>6</sup> in their pioneer investigations, employed methods in which osmotic equilibrium was attained only very slowly. Several days or even two weeks were sometimes required before equilibrium was attained in their experiments. The danger of denaturation or decomposition of the protein under such circumstances is appreciable in many cases, although this danger was reduced to a minimum in Adair's work by carrying out the experiment at  $0^\circ$ . Several more recent workers, however, have devised apparatus

<sup>2</sup> The freezing point depressions of certain proteins in phenol, however, are small enough to show that their

in which equilibrium was attained far more rapidly. Oakley<sup>7</sup> was able to attain osmotic equilibrium in a period which varied between a few hours and one or two days, according to the size of the diffusible impurities. Bourdillon<sup>8</sup> has devised an apparatus requiring only very small amounts of protein solution, in which equilibrium is attained in three to four hours at room temperature. In the apparatus recently employed by Bull<sup>9</sup> equilibrium is attained in two to ten hours. The time taken to achieve equilibrium in an osmotic pressure experiment thus compares not unfavorably with the time required for a sedimentation velocity determination (Chapter 19) and is far shorter than the time generally required in a sedimentation equilibrium study. (3) If the protein is not isoelectric, but is combined with anions or cations, these "bound" ions increase the observed value of the osmotic pressure because of the Donnan equilibria set up<sup>10</sup>. The Donnan effect can be reduced by addition of neutral salts, but is completely absent only in the isoelectric protein. (4) Even when all due precautions are taken, the observed osmotic pressure can give only the mean molecular weight of the non-diffusible components of the system. Thus it cannot be learned, from such measurements alone, whether a given protein preparation is monodisperse or whether it is a mixture of molecules of different sizes. For the present, an answer to this question can be given only by diffusion and sedimentation methods (Chapters 18 and 19).

In spite of all these technical difficulties, Weymouth Reid<sup>11</sup> as early as 1904 obtained fairly satisfactory osmotic pressure measurements on hemoglobin solutions, which gave an estimated molecular weight of about 50,000. Reid's measurements were undoubtedly superior to those of Hüfner and Gansser<sup>12</sup> which suggested a molecular weight near 16,000, although these attracted far more attention at the time than Reid's measurements.

It remained for Sørensen<sup>5</sup> to make a comprehensive and systematic survey of the whole question, and to evaluate correctly the influence of the Donnan equilibrium. On the basis of elaborate and painstaking series of measurements on egg albumin solutions, he estimated the molecular weight of this protein as 34,000. Later studies have led to higher estimates for the molecular weight of this protein, but this great work of Sørensen marked an epoch in the development of the subject, and placed it for the first time on a satisfactory foundation.

A few years later, Adair<sup>6</sup> reinvestigated, with great care and skill, the osmotic pressure of hemoglobin, and deduced a molecular weight near

67,000; that is, four times the minimum molecular weight as determined from the iron content. This value was shortly afterwards confirmed by ultracentrifugal studies<sup>13</sup>.

Since that time, a large number of studies of molecular weight by the osmotic pressure method have been reported by Sørensen, Adair, Burk, Wu, Roche and others. In general, the results from different laboratories have been in good agreement with each other and, for the most part, also with the results obtained by the ultracentrifugal method. In Table 1 are given data on the molecular weight of serum albumin from several different laboratories, compiled, and in some instances recalculated, by Dr. Norval F. Burk; and in Figure 1 a plot of  $\pi/C$  against  $C$  is given for the data from the various laboratories. The data of Sørensen and Burk were obtained at the isoelectric point pH 4.8, those of Adair and Robinson at pH 7.4 where the protein carries a large negative charge. The latter data, therefore, require correction for the effects of unequal distribution of ions across the membrane. Determinations of ion concentration inside and outside and of membrane potentials permit the calculation of the correction factor for the Donnan equilibrium. With this correction, Adair and Robinson showed that a consistent value for molecular weight is obtained in excellent agreement with the data of Sørensen and Burk at the isoelectric point. The data of Roche and Marquet at pH 7.4, though somewhat more irregular, lead also to a mean molecular weight in very good agreement with the others. Data from different laboratories on other proteins have not always given such satisfactory agreement as for serum albumin; nevertheless, there are few outstanding discrepancies. A tabulation of data at present available is given in Tables 2, 3 and 4, which were also compiled for us by Dr. Norval F. Burk.

The observed osmotic pressure at finite concentrations is always greater than that calculated from equations 2 or 3, which are strictly valid only at infinite dilution. An excellent example of the character of the observations found is shown in Fig. 2, taken from Adair's extensive and careful work<sup>13a</sup> on hemoglobin. In this figure, the ratio of the osmotic pressure of protein ions to their concentration in grams per liter is more than three times as great, at the highest concentrations studied, as in a very dilute solution. Even if the concentration of hemoglobin is expressed as grams per thousand grams of solvent rather than grams per liter, the deviation is still large. Many such curves may be described over a wide range of concentrations, by an equation<sup>14</sup> of the form

$$\pi = 100C' \quad (5)$$

TABLE 1. MOLECULAR WEIGHT OF SERUM ALBUMIN FROM MEASUREMENTS OF ITS OSMOTIC PRESSURE

C Concentration per 100 cc. solvent	$\pi$ Observed osmotic pressure	$\phi$ Osmotic coefficient*	$\pi_0$ Corrected osmotic pressure	$\frac{\pi_0}{C}$	$M$ Molecular weight
(a) Solutions at pH 4.8 (isoelectric reaction); buffer, 0.05M acetate; temperature 0°. [Burk (1)]					
(gm)	(cm H <sub>2</sub> O)		(cm H <sub>2</sub> O)		(gm)
0.78	2.39	1.00	2.39	3.06	75,600
1.25	4.01	1.00	4.01	3.20	72,400
2.79	8.53	1.00	8.53	3.06	75,700
2.82	8.68	1.00	8.68	3.08	75,200
3.38	10.80	1.00	10.80	3.19	72,600
3.49	10.91	1.00	10.91	3.13	74,000
4.18	13.01	1.00	13.01	3.11	74,400
8.98	27.84	1.00	27.84	3.10	74,700
12.45	37.69	1.00	37.69	3.02	76,600
Mean.....					74,500
(b) Solutions at pH 7.4; buffer 0.0667 phosphate; temperature 1°. [Adair and Robinson (2)]					
1.08	3.66	1.10	3.34	3.09	75,100
1.50	5.21	1.13	4.59	3.06	75,900
1.52	5.46	1.13	4.83	3.17	73,100
1.60	5.78	1.14	5.07	3.17	75,300
1.62	5.70	1.14	4.98	3.07	75,600
1.89	6.80	1.17	5.81	3.07	75,600
1.97	7.12	1.18	6.05	3.07	75,700
1.99	7.42	1.18	6.31	3.16	73,400
1.99	6.96	1.19	5.87	2.95	78,800
2.00	7.04	1.18	5.94	2.97	78,200
2.08	7.80	1.18	6.61	3.18	73,100
2.11	7.59	1.19	6.36	3.01	77,100
2.23	8.50	1.19	7.13	3.20	72,600
2.48	9.35	1.22	7.68	3.10	75,000
2.54	9.34	1.24	7.56	2.98	78,100
2.88	11.38	1.25	9.09	3.16	73,600
2.89	11.56	1.25	9.26	3.20	72,500
3.06	11.97	1.27	9.45	3.09	75,200
3.20	13.18	1.26	10.42	3.26	71,400
3.64	14.55	1.34	10.89	2.99	77,700
3.72	15.67	1.32	11.86	3.19	72,900
3.76	15.85	1.33	11.95	3.18	73,300
3.92	16.15	1.36	11.91	3.04	76,500
3.99	16.26	1.37	11.87	2.97	78,300
4.77	20.94	1.43	14.66	3.07	75,600
4.81	21.76	1.42	15.37	3.19	72,700
5.04	22.79	1.44	15.78	3.13	74,200
Mean.....					75,100
(c) Solutions at pH 7.4; buffer, 0.0667M phosphate; temperature 0°. [Roche and Marquet (3)]					
0.578	2.04	1.05	1.95	3.37	68,600
0.684	2.35	1.06	2.22	3.25	71,300
1.05	3.56	1.09	3.26	3.10	74,600

TABLE 1.—Continued

C Concentration per 100 cc solvent	$\pi$ Observed osmotic pressure	$\phi$ Osmotic coefficient*	$\pi_0^\dagger$ Corrected osmotic pressure	$\frac{\pi_0}{C}$	$M^\ddagger$ Molecular weight
(d) Solutions at about pH 4.8; solvent, 5 percent (0.74 <i>N</i> ) ammonium sulfate; temperature 0°. [Burk (1) and Sørensen (4)]					
(gm)	(cm H <sub>2</sub> O)		(cm H <sub>2</sub> O)		(gm)
1.88	6.09	1.07	5.67	3.01	76,800
2.47	8.39	1.10	7.66	3.10	74,600
3.90	13.37	1.16	11.54	2.96	78,200
4.13	14.98	1.16	12.94	3.13	73,900
5.06	18.12	1.20	15.12	2.99	77,500
8.58	35.20	1.33	26.38	3.07	75,300
Mean.....					76,000
\$8.98	36.35	1.36	26.69	2.97	77,900
\$13.48	62.80	1.53	41.03	3.04	76,100
\$22.46	129.00	1.88	68.57	3.05	75,900
Mean.....					76,600
Mean of (d).....					76,300
Mean of (a) (b) (c) and (d).....					74,800

\* Calculated from the equation  $\phi = \frac{\pi}{\pi_0} = \frac{\pi}{\pi - BC^2}$ , where  $B$  is a constant equal to the slope of the line obtained by plotting  $\frac{\pi}{C}$  against  $C$ . From Figure 1,  $B = 0$  for (a); 0.276 for (b) and (c); 0.1198 for (d). (Cf. (1) in regard to the derivation of this equation.)

† Calculated from the equation  $\pi_0 = \pi - BC^2 = \frac{\pi}{\phi}$ .

‡ Calculated from the equation,  $M = \frac{RT}{\pi_0} C$ .

§ The measurements of Sørensen, reduced to 0°.

(1) Burk, N. F., *J. Biol. Chem.*, **98**, 353 (1932).

(2) Adair, G. S., and Robinson, M. E., *Biochem. J.*, **24**, 1864 (1930).

(3) Roche, A., and Marquet, F., *Compt. rend. soc. biol.*, **118**, 898 (1935).

(4) Sørensen, S. P. L., "Proteins," The Fleischmann Laboratories, New York, 1925.

in which  $\pi$  is the osmotic pressure,  $C'$  the measured concentration and  $K$  and  $h$  are constants. It has been suggested that  $h$  in this equation may represent a hydration factor giving the amount of water which is bound per gram of protein. It is probable, however, that factors other than hydration are involved in the deviation from a simple linear relation between osmotic pressure of the solute and its weight concentration. The osmotic pressures observed are functions of the activity coefficients of solvent and solute, and these coefficients must be affected by all intermolecular forces, whether these are hydration or other factors. For the present, therefore, we may simply regard equation 5 as an empirical relation which furnishes a good description of the data found for a variety of

hemoglobin over a wide range of concentrations and the relation between osmotic data and activity coefficients has been given by Adair<sup>15</sup>.

A number of proteins have been studied in solvents other than water. Thus hemoglobin<sup>14</sup> and serum albumin<sup>16</sup> have been studied in glycerol and the molecular weight found to be the same in this solvent as in water. Burk and Greenberg demonstrated, however, that the molecular weight in concentrated urea solutions, though for some proteins the same as in water, was for others much smaller. Thus horse hemoglobin, with a molecular weight of 67,000 in water, appears to be reduced to half this size or about 34,000 in urea. Wu and Yang<sup>17</sup> showed that ox hemoglobin

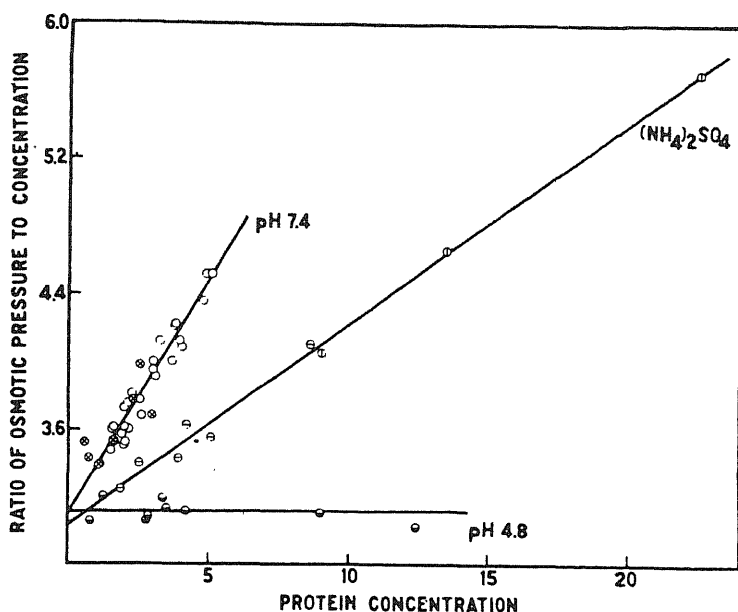


Figure 1

also dissociates in urea but that dog and sheep hemoglobins do not. These results on horse hemoglobin have been confirmed by Steinhardt (Chapter 19), employing the ultracentrifugal method. Proteins can thus be divided into those whose molecular weights are the same in water and urea and those that are somewhat less in urea than in water. The former group includes egg albumin, pepsin, zein, gliadin, and serum albumin. The molecular weights of some of the plant globulins, such as edestin, excelsin and amandin are apparently diminished by an even larger factor in urea

TABLE 2. MOLECULAR WEIGHT OF VARIOUS PROTEINS FROM MEASUREMENTS OF OSMOTIC PRESSURE

Protein	Molecular weight
Egg albumin (1, 2, 3, 4, 5).....	40,000-46,000 <sup>1</sup>
Trypsin (6).....	36,500
Pepsin (7).....	36,000
Chymotrypsin (6).....	41,000
Chymotrypsinogen (6).....	36,000
Zein (8).....	39,000 <sup>2</sup>
Gliadin (9, 10).....	41,000
Hemoglobin (11).....	67,000
.....	73,000 <sup>3</sup>
Serum globulin.....	174,000 <sup>5, 6</sup>
Pseudoglobulin (12).....	177,000
Euglobulin (13).....	174,000
Amandin (10).....	206,000
Excelsin (10).....	214,000
Hemocyanin, king crab ( <i>Limulus</i> ) (14).....	544,000 <sup>7</sup>
Hemocyanin, crab ( <i>Carcinus</i> ) (15).....	550,000 <sup>7</sup>
Hemocyanin, octopus (15).....	710,000
Hemocyanin, snail ( <i>Helix</i> ) (15).....	1,800,000 <sup>7</sup>

The following footnotes apply to Tables 2, 3 and 4.

<sup>1</sup> The study of the molecular weight of egg albumin has had a curious history. Sørensen (5), on the basis of his osmotic pressure measurements, estimated the molecular weight as 34,000, and a practically identical value was found by Burk and Greenberg (16). The earliest studies on the ultracentrifuge [Svedberg and Nichols (27)] indicated a value close to 35,000, and the apparently satisfactory concordance between two independent types of measurement appeared to have settled the question. However, Marrack and Hewitt (3), from osmotic pressure measurements at 37°, estimated the molecular weight of crystalline egg albumin to be 43,000. Adair (1) reexamined Sørensen's data and estimated from them a revised molecular weight of 43,000. Taylor, Adair and Adair (4) gave a value of 46,000 from osmotic pressure measurements, and Bull (2) has recently obtained a value of 45,000 by the same method. Also, within the last few years, the molecular weight of egg albumin has been reexamined by the ultracentrifugal technique and by diffusion measurements in Svedberg's laboratory. The values now given by Svedberg (see Chapter 19, Table 1) are 40,500 from sedimentation equilibrium measurements, and 43,000 from sedimentation velocity and diffusion. The weight of the evidence from physical chemical studies, therefore, points to a molecular weight of  $43,000 \pm 3,000$ .

<sup>2</sup> This value is corrected for deviations from the ideal solution law and for the Donnan effect. Previous values given by Cohn (28) or Burk (10) were uncorrected.

<sup>3</sup> The mean of the values given by Adair and Robinson (23) 72,000; Burk (19) 74,600; Roche and Marquet (24) 69,000; and Sørensen, as corrected by Burk (19) and by Adair and Robinson (23) for measurements in 5 per cent ammonium sulfate solution, 76,300. Because of the different methods employed in calculating the molecular weight, this differs slightly from the mean value given in Table I.

<sup>4</sup> Wu and Yang (20) have studied oxyhemoglobin and methemoglobin of the sheep, dog, and ox in 65% glycerol and have found molecular weights ranging from 64,900 to 65,700 in this solvent for all these proteins. They also report the value 68,000 for ox globin and 62,300 for horse globin in the same solvent. The results of their measurements on the same proteins in 40% urea solution are given in Table 3.

<sup>5</sup> The mean of the values given by Adair and Robinson (23) 175,000 and Burk (12) 173,000, obtained from measurements upon isoelectric solutions (about pH 5.5). Roche and Bracco (26) obtain the value, 150,000, based upon measurements on solutions alkaline to the isoelectric point (pH 7.4) in which a Donnan membrane equilibrium was probably present (cf. (23)).

TABLE 3. MOLECULAR WEIGHTS OF VARIOUS PROTEINS FROM MEASUREMENTS OF THEIR OSMOTIC PRESSURE IN ISOELECTRIC UREA SOLUTIONS

Protein	Molecular weight
Amandin (10).....	30,000
Casein (16).....	33,600
Egg albumin (16, 17).....	34,000
[Pepsin (18).....	36,000 <sup>10</sup> ]
Zein (8).....	37,000 <sup>2</sup>
Gliadin (9, 10).....	44,200
Edestin (16).....	49,000
Serum albumin (19).....	73,800
Serum globulin (12).....	173,000
Hemoglobin (horse) (16).....	34,300
Methemoglobin (sheep) (17).....	66,200
Methemoglobin (dog) (20).....	65,900
Methemoglobin (ox) (20).....	39,200
Oxyhemoglobin (sheep) (20).....	65,600
Oxyhemoglobin (dog) (20).....	65,900
Oxyhemoglobin (ox) (20).....	37,700
Globin (sheep) (20).....	63,000
Globin (dog) (20).....	67,500
Globin (ox) (20).....	34,300

salt solutions acid to the isoelectric point, in which solutions a partial denaturation may have occurred, giving rise to inhomogeneity, his values have less significance than those obtained from measurements at the isoelectric point.

<sup>7</sup> Uncorrected for deviations from the ideal solution law.

<sup>8</sup> This estimate is based upon measurements on the solutions alkaline to the isoelectric point. The elimination of the Donnan effect by extrapolation or by membrane potential measurements involves some uncertainty. Therefore there may not be any real difference in the estimates of this protein from different species, as indicated here.

<sup>9</sup> McFarlane (30) believes that this lower value is due to a contamination of the serum globulin with serum albumin. His preparations of human serum globulin were found to contain 22 to 43% of A fraction (albumin).

<sup>10</sup> This value was derived from the ultracentrifuge experiments of Steinhardt (18), who found that urea caused no dissociation of pepsin.

- (1) Adair, G. S., *Ann. Rev. Biochem.*, **6**, 180 (1937).
- (2) Bull, H. B., *J. Biol. Chem.*, **137**, 143 (1941).
- (3) Marraek, J., and Hewitt, L. F., *Biochem. J.*, **23**, 1079 (1929).
- (4) Taylor, G. L., Adair, G. S., and Adair, M. E., *J. Hygiene*, **32**, 340 (1932).
- (5) Sørensen, S. P. L., *Compt. rend. trav. lab. Carlsberg*, **12**, 262 (1917).
- (6) Kunitz, M., and Northrop, J. H., *J. Gen. Physiol.*, **18**, 455 (1934-35).
- (7) Northrop, J. H., *J. Gen. Physiol.*, **13**, 757 (1929-30).
- (8) Burk, N. F., Unpublished measurements.
- (9) Burk, N. F., *J. Biol. Chem.*, **121**, 46 (1933).
- (10) Burk, N. F., *J. Biol. Chem.*, **120**, 63 (1937).
- (11) Adair, G. S., *Proc. Roy. Soc. (London)*, **A120**, 573 (1928).
- (12) Burk, N. F., *J. Biol. Chem.*, **121**, 373 (1937).
- (13) Adair, G. S., *Skand. Arch. Physiol.*, **49**, 76 (1926).
- (14) Burk, N. F., *J. Biol. Chem.*, **133**, 511 (1940).
- (15) Roche, J., Roche, A., Adair, G. S., and Adair, M. E., *Biochem. J.*, **29**, 2576 (1935).
- (16) Burk, N. F., and Greenberg, D. M., *J. Biol. Chem.*, **121**, 100 (1933).
- (17) Huang, T. C., and Wu, H., *Chinese J. Biol.*, **1**, 1 (1934).
- (18) Steinhardt, J., *J. Biol. Chem.*, **123**, 549 (1939).
- (19) Burk, N. F., *J. Biol. Chem.*, **98**, 353 (1932).
- (20) Wu, H., and Yang, E.-F., *Chinese J. Physiol.*, **6**, 51 (1932).
- (21) Adair, G. S., *J. Am. Chem. Soc.*, **51**, 696 (1929).
- (22) Adair, G. S., *Proc. Camb. Phil. Soc. (Biol.)*, **1**, 75 (1924).
- (23) Adair, G. S., and Robinson, M. E., *Biochem. J.*, **24**, 1864 (1930).
- (24) Roche, A., and Marquet, F., *Compt. rend. soc. biol.*, **118**, 898 (1935).
- (25) Roche, A., Dorier, M., and Marquet, F., *Compt. rend. soc. biol.*, **119**, 1150 (1935).



TABLE 4. MOLECULAR WEIGHT COMPARISONS

(a) Protein	Source	Molec- ular weight
Hemoglobin.....	sheep (21)	68,000
Hemoglobin.....	horse (11)	67,000
Hemoglobin.....	man (22)	66,700
Serum albumin.....	ox (23)	70,000
Serum albumin.....	sheep (23)	70,000
Serum albumin.....	horse (24)	69,000
Serum albumin.....	man (25)	69,000
Serum globulin.....	horse (23)	175,000
Serum globulin.....	ox (23)	174,000
Serum globulin.....	sheep (23)	160,000 <sup>s</sup>
Serum globulin.....	man (26)	103,000 <sup>s</sup>
Pseudoglobulin.....	horse (12)	177,000
Pseudoglobulin.....	sheep (23)	155,000 <sup>s</sup>
Pseudoglobulin.....	ox (23)	150,000 <sup>s</sup>
Euglobulin.....	horse (23)	174,000
Euglobulin.....	sheep (23)	166,000 <sup>s</sup>
(b) Protein	Solvent	Molec- ular weight
Hemoglobin (horse).....	water (dilute buffer) (16)	67,000 <sup>d</sup>
Hemoglobin (horse).....	65% glycerol (16)	66,500 <sup>d</sup>
Serum albumin.....	water (dilute buffer) (19)	74,600
Serum albumin.....	75% glycerol (19)	74,700
Gladin.....	50-60% alcohol, 25-30° (9)	42,000
Gladin.....	15% urethane, 25° (9)	41,000

See notes to table 2.

solutions. Burk has reported the molecular weights of these proteins lie between 30,000 and 50,000 in urea, whereas they are 200,000 to 300,000 in aqueous salt solutions. Weber and Stöver<sup>18</sup> have found an even greater change for myosin, for which they have estimated a molecular weight of the order of magnitude of 1,000,000 in aqueous salt solutions and of approximately 100,000 in urea.

The study of molecular weights as large as 1,000,000 by osmotic pressure is extremely difficult, since the pressures observed are very small. Satisfactory results have, nevertheless, been achieved for some of the hemocyanins<sup>19</sup>. Such large molecules, however, are, in general, much more readily studied by the methods of sedimentation and diffusion which are treated in detail in Chapters 18 and 19.

G. Scatchard, A. C. Batchelder and A. Brown (Buffalo meeting of the

pressure of serum albumin in sodium chloride solutions, over a wide range of pH without buffer, and have also measured the distribution of sodium chloride across the membrane. They find it convenient to express the concentrations as moles per kilogram of water and to define the protein component as  $\text{Na}_{z/2} \text{AlbCl}_{z/2}$ , in which Alb represents a protein ion with

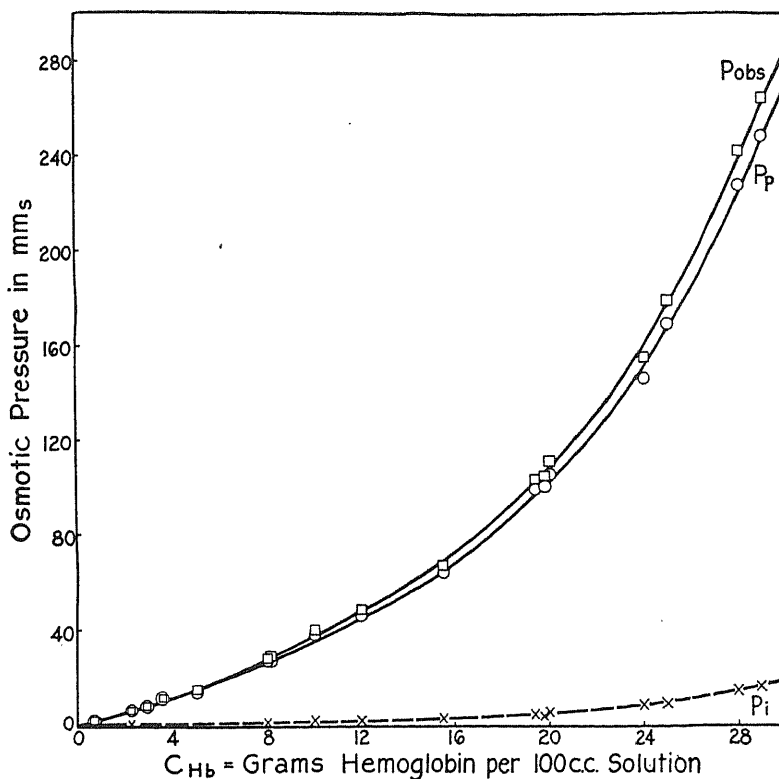


Figure 2.  $P_{obs}$  = observed osmotic pressure of hemoglobin in millimeters of mercury at  $0^\circ$ .  $P_p$  = partial pressure of the protein ions.  $P_i$  = diffusible ion pressure difference calculated by formula 34 of Adair [Proc. Roy. Soc. London, A120, 595 (1928)].  $C_{Hb}$  = grams dry hemoglobin per 100 cc. protein solution, in equilibrium with phosphate buffer mixture at pH 7.8. Data from Adair, G. S., Proc. Roy. Soc. London, A120, 595 (1928).

valence  $z$ . This gives an electrically neutral component with one molecule for each protein ion, and also corresponds to the continuous changes in properties as  $z$  changes from positive to negative. The third component is sodium chloride whose concentration,  $m$ , is the average of the sodium

outside,  $m_3'$ , is proportional to the protein concentration  $m_2$  within the error of measurement. The ratio of the osmotic pressure to the protein concentration,  $\pi/m_2$ , is an almost linear function of  $m_2$  even up to 300 g albumin per kilogram of water, and a straight line was assumed up to 60 g per kilo.

The extrapolated value of  $\pi/m_2$  is independent of the valence of the protein and corresponds to a molecular weight of 69,000. The slope has a very flat minimum when  $z$  is about  $-20$ .

The activity coefficients are related to these measurements by the equations

$$\ln \frac{m_3}{m_3'} = -\frac{\beta_{23}}{1 + \beta_{33} m_3/2}$$

$$\frac{\pi}{m_2} = \frac{RT}{V_m} \left[ 1 + m_2 \left( \frac{z^2}{4m_3} + \frac{\beta_{22}}{2} - \frac{\beta_{23}^2 m_3/4}{1 + \beta_{33} m_3/2} \right) + \dots \right]$$

in which  $V_m$  is the volume of the *outside* solution containing one kilo of solvent, the first term in the parenthesis represents the Donnan effect for ideal solutions,

$$\beta_{22} = \frac{\partial \ln \gamma_2}{\partial m_2}, \quad \beta_{23} = \frac{\partial \ln \gamma_2}{\partial m_3} = \frac{2\partial \ln \gamma_3}{\partial m_2}, \quad \text{and} \quad \beta_{33} = \frac{2\partial \ln \gamma_3}{\partial m_3}$$

These are all limits at zero salt concentration. The first two  $\beta$ 's appear to vary but little with protein concentration, but all change with the salt concentration. Since  $\beta_{33}$  is determined by measurements of the activity of sodium chloride in the absence of protein,  $\beta_{23}$  may be determined from the salt distribution and then  $\beta_{22}$  may be determined from the osmotic pressure.

In 0.15 *m* NaCl,  $\beta_{23}$  is nearly constant at  $-21$  from  $z$  equal to  $-25$  to  $+10$ , but it rapidly becomes larger for more positive values of  $z$ . For isoionic ( $z = 0$ ) horse serum albumin and  $\text{CaCl}_2$ , Joseph's<sup>20</sup> electromotive force measurements lead to  $-23 \pm 4$ , but there is no reason to expect the two salts to give the same constant within a factor of two. Particularly in dilute salt solutions the variation of  $\beta_{23}$  with salt concentration is more rapid than the inverse square root relation used empirically by Joseph.

On the other hand,  $\beta_{22}$  varies but little with the salt concentration, but changes rapidly with the valence. In 0.15 *m* NaCl it is positive from  $z$  equal to  $-23$  to  $+11$  with a maximum of about 900 when  $z$  is about 5; it is  $-2000$  when  $z$  is 27.

The explanation of these activity coefficients must be complicated, and there is danger in attributing the whole effect. even at the isoionic point.

to a single factor such as hydration. Calculated thus from  $\beta_{22}$  the hydration would be large and positive, but calculated from  $\beta_{23}$  it would be negative. Those parts of these coefficients which vary as the square of the valence may be attributed to electrostatic effects depending upon the net charge, but there is no simple way to distinguish electrostatic effects which depend upon the dissymmetry of the charge from those effects which are also operative with uncharged molecules.

## Chapter 18

### Translational Diffusion of Amino Acids and Proteins

BY JOHN T. EDSALL AND JOHN W. MEHL

In a state of equilibrium in the absence of external forces, the concentration of any molecular species, at constant temperature and pressure, is uniform throughout any single phase. If the distribution of molecules is not uniform, they will tend to move from regions of higher to those of lower concentration. Such a flow is a necessary consequence of the second law of thermodynamics, since the entropy of the system is at a maximum when the molecules are distributed with statistical uniformity throughout. This process of flow, or diffusion, results from the thermal energy of the molecules, which gives rise to their Brownian movement. The speed with which a given molecule diffuses, under specific conditions, is indicated by its diffusion constant; and the latter is a function of the size and shape of the molecule.

The Brownian movement of molecules (on which their diffusion depends) involves not only translation but also rotation. To consider the effect of the latter factor, let us imagine a system of ellipsoidal molecules, in solution, all alike, whose orientation may be described in terms of the angles between their principal axes and some fixed direction in the surrounding medium. If the molecules are oriented by an external force, so that their principal axes are all parallel, and this orienting force is then suddenly removed, the parallel orientation will gradually disappear. In the final state of equilibrium, the distribution of orientations among the molecules will be completely random<sup>1</sup>. The rotary diffusion constant is a measure of the speed with which this equilibrium state is approached. Inversely proportional to the rotary diffusion constant is the relaxation time, which is proportional to the time required for the molecules to assume a random orientation. A spherical molecule may be characterized in terms of a single rotary diffusion constant (or relaxation time). An ellipsoid of revolution will require two such constants for its characterization; one for motion about the axis of revolution, and one for motion about the

The study of translational and rotary diffusion provides the most powerful set of physical methods at present available for determining the size and shape of large molecules in solution. In the determination of the translational diffusion constant, free diffusion is permitted to take place. Even more important in many cases, however, is the study of diffusion modified by the presence of an external force, producing a directed flow of solute. Thus, in the determination of molecular weight by sedimentation equilibrium, as developed by Svedberg, the tendency of the molecules to diffuse freely is opposed by a centrifugal field, and the final state of equilibrium is determined by the balance between these two opposing forces. In the study of sedimentation velocity in the ultracentrifuge (Chapter 19), sedimentation is opposed by the frictional resistance of the medium. The latter is inversely proportional to the diffusion constant of the molecules, which must therefore be evaluated to obtain the molecular weight from sedimentation data.

Similarly, in the determination of rotary diffusion constants, the Brownian movement of the molecules is opposed by an external constraining force, which tends to orient them. This orienting force may be a shearing stress produced by a velocity gradient in the liquid, as in studies of viscosity and double refraction of flow (Chapter 21). If the molecules are electric dipoles, the external force may be an electric field, as in studies of the dispersion of the dielectric constant (Chapter 22).

All these methods, and the results obtained from them, will be discussed in detail in later chapters. In the present chapter, the laws of diffusion, and the relation of diffusion constants to molecular size and shape, will be set forth.

### Translational Diffusion

In 1855, Adolf Fick<sup>2</sup> suggested that the process of diffusion could be treated as analogous to that of the conduction of heat. Thus, if diffusion is taking place in the  $x$  direction across an area  $Q$ , the mass of solute  $dS$  diffusing in the time  $dt$  at any point should be proportional to the concentration gradient  $\frac{dc}{dx}$  at that point, and to the area  $Q$ .

$$dS = -D \cdot Q \cdot \frac{dc}{dx} dt \quad (1)$$

The minus sign denotes simply that the solute diffuses in the direction of decreasing concentration.

This relation is known as Fick's first law; and  $D$  in equation 1 is the

presented subsequently will be expressed in  $\text{cm}^2/\text{sec}$ . For the amino acids,  $D$  in water at  $20^\circ$  is of the order of  $10^{-5} \text{ cm}^2/\text{sec}$ ; for the proteins, under the same conditions, values of  $D$  between  $10^{-6}$  and  $10^{-7} \text{ cm}^2/\text{sec}$  have been found.

The change of concentration with time at any point along the diffusion column is given by the relation

$$\frac{\partial c}{\partial t} = D \left[ \frac{\partial^2 c}{\partial x^2} + \frac{1}{Q} \frac{dQ}{dx} \cdot \frac{\partial c}{\partial x} \right] \quad (2)$$

If  $Q$  is constant throughout the diffusion column, equation 2 becomes

$$\frac{\partial c}{\partial t} = D \frac{\partial^2 c}{\partial x^2} \quad (3)$$

which is Fick's second law<sup>3</sup>.

These equations imply that  $D$  is independent of the concentration. In general, this is not true, but the analysis of diffusion experiments is usually made on the assumption that  $D$  can be treated as a constant over a moderately wide range of concentrations. When this is not true, the experimental values obtained are mean values over a given concentration range. By determining  $D$  over several ranges, approaching zero concentration, an extrapolated value of  $D$  for infinite dilution can be obtained, and it is this value which is commonly most important for the theoretical interpretation of the data.

We may now consider two important special cases of systems to which Fick's law is applicable, assuming that the solutions involved are so dilute that  $D$  is independent of the concentration. In these cases the diffusion is assumed not to be restricted by the dimensions of the cell in which the process takes place; that is, the diffusion column is long enough and the time is short enough so that the concentration changes do not occur at the extreme ends of the column.

**Case I: Diffusion from Solution to Solvent.** At the beginning of the experiment (time  $t = 0$ ) the concentration of the solute in the negative half of the cell is  $c_0$ ; in the positive half it is zero. Then analysis shows<sup>3</sup> that at any time after diffusion begins the concentration of solute as a function of time and distance is given by the equation

$$c = \left\{ \frac{c_0}{2} \{1 - P(z)\} \right\} \quad (4)$$

where  $z = \frac{x}{\sqrt{4Dt}}$ , and  $P(z)$ , the probability integral, is

The form of the concentration function for various times after diffusion has started is shown in Fig. 1.

The concentration gradient,  $\frac{\partial c}{\partial x}$ , at any point and time is

$$\frac{\partial c}{\partial x} = \frac{-c_0}{\sqrt{4\pi Dt}} e^{-x^2/4Dt} \quad (6)$$

that is, this gradient, plotted as a function of  $x$  at any given time, has the form of an ideal Gaussian distribution curve (see Fig. 2); and the slope of this gradient at any point is

$$\frac{\partial^2 c}{\partial x^2} = \frac{c_0 x}{4\pi^{1/2} D^{3/2} t^{3/2}} e^{-x^2/4Dt} \quad (7)$$

The rate at which the concentration is changing with time at any given value of  $x$  is

$$\frac{\partial c}{\partial t} = \frac{c_0 x}{4\pi^{1/2} D^{1/2} t^{3/2}} e^{-x^2/4Dt} = D \frac{\partial^2 c}{\partial x^2} \quad (8)$$

showing that the given distribution satisfies Fick's second law (equation 3).

**Case II: Diffusion from a Thin Layer.** At time  $t=0$  the concentration of the solute in the cylinder is zero except for an infinitely thin layer of thickness  $dx$  between the positive and negative halves of the vessel, in which it becomes infinite as the thickness approaches zero. Under these conditions (see ref. 3) the concentration function at any time after the beginning of the experiment is

$$c = \frac{1}{\sqrt{4\pi Dt}} e^{-x^2/4Dt} \quad (9)$$

Thus the concentration of solute in this system (Fig. 2) as a function of time and distance is given by the same equation that describes the concentration gradient in the system first considered (equation 6).

In 1888, Nernst<sup>4</sup> attempted to examine the process of diffusion more intimately, and to relate it to forces active in osmosis and in electrical conductance and transference. For the case of solutions of non-electrolytes which are sufficiently dilute that the osmotic pressure may be considered to be proportional to the concentration

$$\pi = RTC \quad (10)$$

he concluded that the diffusion coefficient should be given by

$$D = RT/F \quad (11)$$

1. If the force which must act on one gram mole of the diffusing sub-



to the case of a uni-univalent electrolyte, he pointed out that the quantity  $F$  for each ion may be evaluated from the electrical mobility, which had been determined by Kohlrausch from conductance measurements, in very dilute solutions. Nernst also showed that the difference in mobility of the

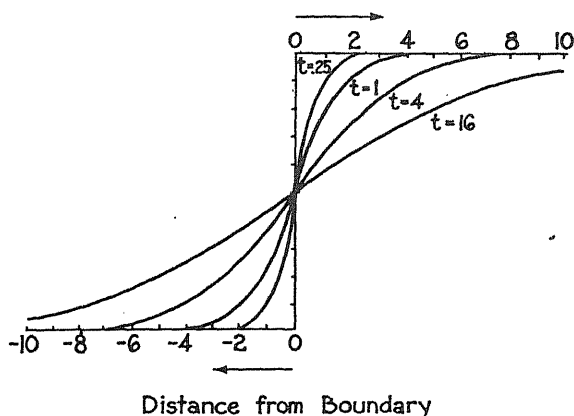


Figure 1. Relative concentration as a function of distance from boundary. Diffusion from solution across boundary into solvent. From Williams, J. W. and Cady, L. C., *Chem. Rev.*, 14, 171 (1934).

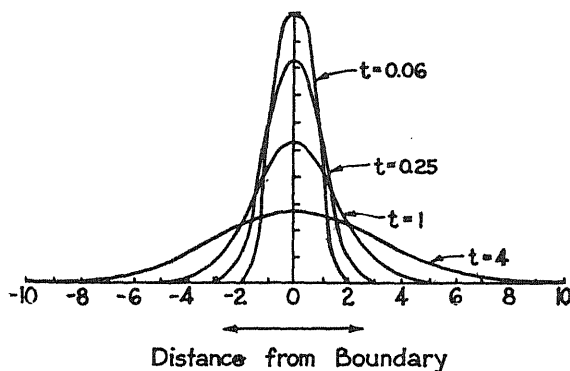


Figure 2. Relative concentration as a function of distance from boundary. Diffusion from thin layer into solvent. From Williams, J. W. and Cady, L. C., *Chem. Rev.*, 14, 171 (1934).

positive and negative ions would tend to separate them during diffusion, and that the consequence of this tendency to separate would be the estab-

move under the osmotic force alone. The potential gradient for a uni-univalent electrolyte is given by the expression ( $P$  = potential)

$$\frac{dP}{dx} = \frac{1}{c} \left( \frac{u_c - u_a}{u_c + u_a} \right) \frac{d\pi}{dx} \quad (12)$$

where  $u_a$  and  $u_c$  are the mobilities of anion and cation, and  $\pi$  is, as before, the osmotic pressure. In terms of the limiting equivalent conductance of anion and cation, the value of  $D$  at infinite dilution is

$$D_0 = \frac{2RT}{9.31 \times 10^{16}} \frac{\Delta_a - \Delta_c}{\Delta_a + \Delta_c} \quad (13)$$

From the considerable mass of diffusion data then at hand, Nernst calculated values of  $F$  (equation 11) for a number of non-electrolytes, but concluded that it was not yet possible to decide whether  $F$  could be related to the weight, constitution and configuration of the molecule concerned. Nevertheless the values obtained did show a regular increase with increasing molecular weight.

It is evident that these theoretical relations are limited in their application. The treatment of the driving force in diffusion as directly deducible from osmotic pressure data is valid only at infinite dilution, for the driving force in diffusion must be the escaping tendency of the molecular or ionic species involved, whereas osmotic pressure measurements give the escaping tendency of the solvent. In addition to the approximation in the formulation, changes in the escaping tendency with concentration may be expected to lead to changes in  $D$  as the concentration changes. Changes in  $F$  with concentration will also lead to changes in  $D$ . The changes in  $F$  for ions cannot be inferred directly from changes in electrical conductance, for in the latter case positive and negative ions are moving in opposite directions, while in diffusion they are moving together; but there can be no doubt that the interactions between ions and molecules in concentrated solutions will lead to some change in mobility. An attempt has been made to formulate the expression for diffusion more exactly by Onsager and Fuoss<sup>5</sup> and they have obtained the expression

$$D = \frac{\Omega}{C} RT \left[ 1 + C \frac{\partial \ln \gamma}{\partial C} \right] \quad (14)$$

where  $\Omega/C$  is a "mobility"—that is, the velocity acquired under the action of unit force, by the molecular species involved—and the other terms have their usual significance. An attempt was also made by them to evaluate  $\Omega/C$  as a function of the concentration for electrolytes and to apply equa-

some inadequacy in the theory, or to errors in the experimental data, the agreement was disappointing. An unsatisfactory attempt was also made by Mehl and Schmidt<sup>6</sup> to apply equation 14 to data for the diffusion of amino acids. The assumption made in this case, that  $\Omega/C$  could be considered independent of the concentration, was undoubtedly at fault. Gordon<sup>7</sup> has employed an equation, closely related to that of Onsager and Fuoss:

$$D = D_{\infty} \left( 1 + C \frac{\partial \ln \gamma}{\partial C} \right) \left( \frac{\eta_0}{\eta} \right) \quad (14a)$$

in which  $D_{\infty}$  is the diffusion constant of the solute at infinite dilution, and  $\eta_0$  and  $\eta$  are the viscosities of the pure solvent and of the solution, respectively. Gordon and his collaborators<sup>8</sup> have shown that this equation fits the data for several strong electrolytes very well indeed.

**Diffusion and Brownian Movement.** In 1905, William Sutherland<sup>9</sup> published "A Dynamical Theory of Diffusion for Non-Electrolytes, and the Molecular Mass of Albumins," in which he applied Stokes' law to the calculation of the force,  $F$ , of equation 11. For the case of large spherical molecules, where there is no slip between the diffusing molecule and the solvent molecules, he obtained the relation

$$D = \frac{kT}{6\pi\eta r} = \frac{RT}{6\pi\eta rN} \quad (15)$$

where  $\eta$  is the viscosity of the solvent and  $r$  is the radius of the diffusing molecule<sup>10</sup>. Thus the diffusion constant is directly proportional to the mean kinetic energy of the diffusing molecule; and inversely proportional to the frictional resistance it encounters in its motion; since from Stokes' law the resisting force which a sphere moving through a viscous medium with a small constant velocity  $v$  must overcome is

$$\phi = 6\pi\eta r v \quad (16)$$

It is of some historical interest that Sutherland applied equation 15 to the data of Graham which had been analyzed by Stefan<sup>11</sup> and obtained a molecular weight of 33,000 for egg albumin.

In the following year, the phenomenon of the Brownian movement was for the first time given a satisfactory theoretical interpretation by Einstein<sup>12</sup>

<sup>6</sup> Mehl, J. W. and Schmidt, C. F., *Univ. Cal. Publ. Physiol.*, 8, No. 13 (1937).

<sup>7</sup> Gordon, A. R., *ibid.*, 5, 522 (1937).

<sup>8</sup> (a) James, W. A., and Gordon, A. R., *J. Chem. Physics*, 7, 89 (1939); (b) James, W. A., and Gordon, A. R., *ibid.*, 8, 423 (1940).

<sup>9</sup> Sutherland, W., *Philosophical Magazine*, 6, 799 (1905).

<sup>10</sup> (a) Mehl, J. W., and Schmidt, C. F., *Univ. Cal. Publ. Physiol.*, 8, No. 13 (1937).

<sup>11</sup> Stefan, J., *Ann. Phys.*, 178, 269 (1855).

<sup>12</sup> Einstein, A., *Ann. Phys.*, 17, 549 (1905).

and Smoluchowski<sup>13</sup>, and a far deeper insight into the molecular basis of diffusion was thereby obtained. Only a very simple outline of their reasoning need here be given. The thermal motion of a molecule is to be regarded as purely random; considering only the component of the motion along one direction—the  $x$  axis—it is therefore apparent that positive and negative displacements are equally probable. The probability of a net horizontal displacement having a value between  $x$  and  $x + dx$  is

$$P(x) dx = \frac{1}{\sqrt{2\pi\Delta^2}} e^{-x^2/2\Delta^2} dx \quad (17)$$

where  $\Delta^2$  is the mean square of the displacement in the  $x$  direction. This mean square value may be found by determining the net displacements of a very large number of like molecules over the same time interval; or else by determining the net displacement of a single molecule in a large succession of repeated time intervals, all intervals being of the same length<sup>14</sup>.

It has been shown by Einstein that  $\Delta^2$  bears a very simple relation to the diffusion constant of the molecule. In its simplest form the argument is as follows: Imagine the diffusion as taking place across a plane perpendicular to the axis of a uniform tube of unit cross-section. Let the concentrations to the left and right of this plane be  $c_1$  and  $c_2$  respectively. The mean number of molecules passing from left to right across this plane (from a mean distance— $\Delta/2$  on the left to a mean distance +  $\Delta/2$  on the right) in time  $t$  is

$$n_1 = \frac{c_1 \Delta}{2} \quad (18)$$

The mean number passing in the opposite direction in the same time is

$$n_2 = \frac{c_2 \Delta}{2} \quad (19)$$

and the net transference from left to right is

$$n_1 - n_2 = (c_1 - c_2) \frac{\Delta}{2} \quad (20)$$

If  $\Delta$  is small

$$\frac{c_2 - c_1}{\Delta} = \frac{dc}{dx}$$

and the net number of molecules passing from left to right in unit time is

$$\frac{n_1 - n_2}{t} = \frac{-\Delta^2 dc}{2t dx} \quad (21)$$

But by Fick's first law, the diffusion constant  $D$  is the number of molecules passing unit cross-section in unit time when the concentration gradient  $-dc/dx$  is unity; therefore

$$D = \frac{1}{2} \frac{\Delta^2}{t} \quad (22)$$

and  $P(x)$  in equation 17 becomes

$$P(x) = \frac{1}{\sqrt{4\pi Dt}} e^{-x^2/4Dt} \quad (23)$$

and this (see equation 9) is a solution of Fick's second law.

Einstein also generalized the reasoning of Nernst and Sutherland, showing  $D$  to be a function of the dimensions of the molecules and the viscosity of the liquid. From equation 11,

$$D = \frac{p_0}{F} = \frac{RT}{F} \text{ per gram mole} \quad (24)$$

$$= \frac{kT}{f} \text{ per molecule} \quad (24a)$$

where the relation  $p_0 = RT$  follows from van't Hoff's law of osmotic pressure in very dilute solutions, and  $f$  is the force which must act on a molecule to give it a velocity of 1 cm/sec in the medium of viscosity  $\eta$ . For a spherical particle,  $f = 6\pi\eta r$  from equation 16, and the diffusion constant is therefore given by equation 15.

For non-spherical molecules, the frictional coefficient  $F$  cannot be evaluated from this simple formula. If the molecule is an ellipsoid, then the resistance to its motion is characterized by three different frictional coefficients,  $f_1, f_2, f_3$ , each coefficient characterizing the resistance to motion of the ellipsoid parallel to one of its principal axes; and there are three corresponding diffusion constants.

$$D_1 = \frac{kT}{f_1} \quad (25a); \quad D_2 = \frac{kT}{f_2} \quad (25b); \quad D_3 = \frac{kT}{f_3} \quad (25c)$$

The actual free diffusion of a system of such molecules, however, does not follow a direction parallel to any of these axes; the molecules are oriented at random and are diffusing in the  $x$  direction. Under these circumstances, the experimentally determined diffusion constant  $D$  is

$$D = \frac{kT}{f} \quad \text{where } \frac{1}{f} = \frac{1}{3} \left( \frac{1}{f_1} + \frac{1}{f_2} + \frac{1}{f_3} \right)$$

For an ellipsoid of revolution,  $f_2 = f_3$  in the above formulas, and the numerical evaluation of the diffusion constant as a function of the dimensions of the ellipsoid becomes possible. Let  $b$  be the length of the equatorial semi-axis of the ellipsoid,  $a$  the length of the semi-axis of revolution, and  $\rho$  the ratio  $b/a$ . Perrin has shown that the diffusion constant is given by the formula

$$\frac{D_0}{D} = \frac{f}{f_0} = \frac{(1 - \rho^2)^{1/2}}{\rho^{2/3} \ln \frac{1 + (1 - \rho^2)^{1/2}}{\rho}} \quad (27)$$

for an elongated ellipsoid ( $\rho < 1$ ), and by

$$\frac{D_0}{D} = \frac{f}{f_0} = \frac{(\rho^2 - 1)^{1/2}}{\rho^{2/3} \tan^{-1} (\rho^2 - 1)^{1/2}} \quad (28)$$

for a flattened ellipsoid ( $\rho > 1$ )<sup>16</sup>

$D_0$  in these equations is the diffusion constant of a sphere of the same mass and volume as the ellipsoid; it can be evaluated from equation 15 if the molecular weight ( $M$ ) and the partial specific volume ( $V$ ) of the protein and the viscosity ( $\eta$ ) of the solvent, are known.

$$D_0 = \frac{kT}{f_0} = \frac{kT}{6\pi\eta \left( \frac{3VM}{4\pi N} \right)^{1/3}} \quad (15a)$$

$f/f_0$  is the "frictional ratio" of Svedberg and Pedersen (17). Values of  $a/b$  for various values of  $D_0/D$ , calculated from equations 27 and 28, are given in Table 1, taken from Svedberg and Pedersen<sup>17, p. 41</sup>.

Equations 27 and 28 permit estimations of the shape of protein molecules if their molecular weights, densities, and diffusion constants are known. The results so obtained are set forth in Chapter 19. The significance of the molecular shapes evaluated from these equations depends on the assumption that the protein molecules can be regarded approximately as ellipsoids of revolution. The extent to which such an assumption is justified is still an open question. All these equations assume that the diffusing molecule is very large in comparison with the solvent—an assumption probably justified in the case of the proteins—and that there is no slip between the diffusing molecule and the solvent molecules. Also the degree of hydration of the diffusing protein molecules raises a difficult and still unsettled problem. The molecular weights and densities used in the evaluation of  $D_0$  in equations 27 and 28 are generally values for the anhydrous protein. If the diffusing protein carries attached to it a certain

solvation effect. The magnitude of the solvation is uncertain. The water of hydration of various protein crystals has been estimated by Adair and Adair<sup>18</sup>, who have given values ranging between 0.22 and 0.34 gram of water per gram of protein (Chapter 16). If a diffusing protein molecule carries with it 0.3 gram of water per gram of protein, then the molecular volume of the hydrated protein is approximately 40% greater than that of the anhydrous protein (allowing for the difference in specific volume between protein and water) and the hydration should diminish  $D_0$  by approximately 10%. As will be seen in the next chapter, however, the observed values of  $D/D_0$  are often much lower than can be accounted for by this

TABLE 1. THE AXIAL RATIO OF ELLIPSOIDAL MOLECULES AS A FUNCTION OF THE FRICTIONAL RATIO,  $f/f_0$

Elongated ellipsoids (equation 27)				Flattened ellipsoids (equation 28)			
$\frac{1}{\rho} = \frac{a}{b}$	$\frac{f}{f_0} = \frac{D_0}{D}$	$\frac{1}{\rho} = \frac{a}{b}$	$\frac{f}{f_0} = \frac{D_0}{D}$	$\rho = \frac{b}{a}$	$\frac{f}{f_0} = \frac{D_0}{D}$	$\rho = \frac{b}{a}$	$\frac{f}{f_0} = \frac{D_0}{D}$
1.0	1.000	12	1.645	1.0	1.000	12	1.534
1.2	1.003	14	1.739	1.2	1.003	14	1.604
1.4	1.010	16	1.829	1.4	1.010	16	1.667
1.6	1.020	20	1.996	1.6	1.019	20	1.782
1.8	1.031	25	2.183	1.8	1.030	25	1.908
2.0	1.044	30	2.356	2.0	1.042	30	2.020
3.0	1.112	35	2.518	3.0	1.105	35	2.119
4.0	1.182	40	2.668	4.0	1.165	40	2.212
5.0	1.255	50	2.946	5.0	1.224	50	2.375
6.0	1.314	60	3.201	6.0	1.277	60	2.518
7.0	1.375	70	3.438	7.0	1.326	70	2.648
8.0	1.433	80	3.658	8.0	1.374	80	2.765
9.0	1.490	90	3.867	9.0	1.416	90	2.873
10.0	1.543	100	4.067	10.0	1.458	100	2.974

The values given in this table are for unhydrated molecules. The effect of hydration on the calculated axial ratio is discussed in the text.

From Svedberg, T. and Pedersen, K. O., p. 41, "The Ultracentrifuge," Clarendon Press, Oxford, 1940.

calculation; and most of the effects observed must therefore be attributed to asymmetry of the protein molecule rather than to hydration. Available data on the rotary diffusion constants of protein molecules (Chapters 21 and 22) also give important evidence in favor of this conclusion.

### Experimental Methods

There are two important types of experimental methods in use for the study of the diffusion of the proteins and amino acids. The first is of the classical sort, involving free diffusion from a solution of the molecule being studied into a layer of solvent above the solution. The second method in-

The classical method has been considerably refined by workers at the Institute of Chemistry and Physical Chemistry at the University of Upsala. In this type of experiment it is essential that the boundary initially formed between the solution and solvent be very sharp, and that a very accurate temperature control and freedom from mechanical vibration be maintained. In addition, it is highly desirable to use analytical methods which do not necessitate sampling and the disturbances involved in such a procedure; this is best achieved by optical methods of analysis. In the method of Tiselius and Gross<sup>19</sup> the concentration of the protein in the diffusion column is determined by photographing the solution, using light of a wave-length absorbed by the protein but not by the solvent. From the blackening of the photographic plate the concentration of protein throughout the diffusion column may be calculated; a series of exposures is taken in successive time intervals. The concentration of protein, as a function of distance from the initial boundary, is given by equation 4, provided that the protein molecules are uniform in size and shape, and that the diffusion constant is independent of the concentration.

A second optical method is that of Lamm<sup>20</sup> and Lamm and Polson<sup>21</sup> and is based on the refraction of light when it passes through a medium of varying index of refraction. Before the diffusion study is made, the protein solution is dialyzed at low temperature, usually for several days, against an outer liquid of exactly known composition, and the outside solvent is replaced and renewed from time to time, in order to insure that the composition of the solvent (and hence the refractive index) is determined. A uniform, transparent scale placed behind the vertical diffusion cell is photographed through the cell, and as the concentration gradient changes during diffusion the photographed scale lines become displaced as compared with the normal scale. The measurements yield a relation between the *concentration gradient* and the distance from the boundary, provided the refractive index is a linear function of concentration. This gradient is given by equation 6, which expressed in terms of refractive index of the solution becomes (see Lamm<sup>20</sup>):

$$\frac{\partial n}{\partial x} = \frac{n_1 - n_0}{\sqrt{4\pi Dt}} e^{-x^2/4Dt} \quad (29)$$

where  $n_0$  and  $n_1$  are respectively the index of refraction of solvent and protein solution. If the concentration gradient at any time be plotted as a function of distance ( $x$ ) from the boundary, an ideal Gaussian distribution curve (Fig. 2) should be obtained if the protein molecules are uniform in size and shape and the diffusion constant is independent of concentration.



simple law of free diffusion, may thus be detected by deviations between the ideal and the experimental distribution curve. Unfortunately, however, diffusion measurements are very insensitive as an indication of polydispersity, which is much more readily detected by sedimentation measurements.

This method has also been used by Neurath<sup>22, 23, 24</sup> in a number of investigations; by Entrikin<sup>25</sup> and others in the laboratory of J. W. Williams at the University of Wisconsin; and by Mehl in unpublished investigations from this laboratory.

The use of the porous disk was introduced by Northrop and Anson<sup>26</sup>, and has been considerably extended by McBain and co-workers. This method has the advantage that precautions taken for accurate temperature control and the elimination of mechanical vibration need not be so elaborate. It has, however, the disadvantage that the pores of the disk may become clogged with air bubbles or with solids, that substances may be adsorbed in the material of the disk, that the pores may be small enough to hinder the passage of very large molecules, and that some reference value must be used for calibration.

The problem of calibrating the sintered glass diffusion cell has actually been one of the most troublesome which has had to be met. Northrop and Anson<sup>26</sup> had originally used hydrochloric acid for the calibration, taking an extrapolated value from the free diffusion data of Ohlm. McBain and Liu<sup>27</sup> preferred to use potassium chloride for the calibration, taking the diffusion coefficient of a 0.1*N* solution as determined by Cohen and Bruins<sup>28</sup>. The use of either of these standards is, however, open to certain objections. For both of these substances the diffusion coefficient changes rapidly with the concentration in the region below 0.1*N*. It has already been noted that the diffusion coefficient obtained experimentally is an average value over a range of concentrations. Although the concentration range may be the same for free diffusion measurements and for the disk method, the way in which the concentration varies along the diffusion column will differ in the two cases, and the two types of experiment will not be comparable. In the application of equations 1 or 3 to the analysis of diffusion experiments, they must be integrated according to the experimental conditions. If *D* does not vary with the concentration, or if account of its variation can be taken, the results should be independent of the method. When, however, the integration is made on the assumption that *D* is independent of the concentration, the relation between the values

<sup>22</sup> Neurath, H., *Cold Spring Harbor Symp. Quant. Biol.*, 6, 196 (1938).

<sup>23</sup> Neurath, H., and Cooper, G. R., *J. Biol. Chem.*, 135, 455 (1940).

<sup>24</sup> Neurath, H., Cooper, G. R., and Entrikin, J. O., *J. Biol. Chem.*, 139, 411 (1941).

of the mean diffusion coefficients obtained from two types of diffusion experiment will depend upon the way in which  $D$  varies with the concentration. For this reason it would be much more satisfactory to use for calibration a substance which has a diffusion coefficient independent of the concentration. Northrop and Anson<sup>29</sup> have recently proposed the use of concentrated sodium chloride solutions. Between concentrations of about 0.2 to 2.0*N* the change in the diffusion coefficient is relatively small, and values from the disk and free diffusion methods might be expected to be directly comparable.

Mehl and Schmidt<sup>6</sup> have also considered the problem of the calibration of the sintered glass diffusion cell, and obtained values in agreement with the sodium chloride calibration of Northrop and Anson by two other methods which did not involve the use of other diffusion data. However, the value obtained for the diffusion coefficient of 0.1*N* potassium chloride was, on this basis, about 12% lower than the value obtained by Cohen and Bruins (confirmed by Lamm<sup>20</sup>) by a free diffusion method. For the reasons which have been discussed, this was not entirely unexpected, but values obtained by Mehl and Schmidt for the diffusion coefficients of a series of amino acids are also about 12% lower than those obtained by Polson<sup>30</sup>. Since the diffusion coefficients of the amino acids are not markedly dependent upon the concentration it might be reasonably expected that the two methods should agree in this case.

Important studies, dealing with the calibration of the sintered glass diffusion cell, have recently been published by James, Hollingshead and Gordon<sup>8a</sup> and by Hartley and Runnicles<sup>31</sup>. If we consider the diffusion of solute across a porous membrane from the concentration  $c'$  to concentration  $c''$ , then an apparent diffusion constant,  $D'$ , may be computed from the observed amount of solute diffusing per second.

$$D'(c' - c'') = \int_{c''}^{c'} D \, dc \quad (30)$$

where the true diffusion constant,  $D$ , is a function of  $c$ . Then an effective diffusion constant,  $\bar{D}$ , may be defined by the relation

$$\bar{D}\theta = \int_0^\theta K' \, dt \quad (31)$$

where  $\theta$  is the time elapsed, from the beginning of the measurement<sup>7, 8a</sup>. These relations hold for any value of  $c'$  and  $c''$ , provided that the concentration gradient within the diffusion membrane is at all times that for a steady state.

James, Hollingshead and Gordon<sup>8a</sup> have chosen as the best value of  $\bar{D}/D_{\infty}$  for the diffusion of 0.1*N* KCl into water, 0.924 at 25°. This corresponds to 0.925 at 18°. Hartley and Runnicles<sup>31</sup>, by an entirely different method, arrive at the value 0.922 at 18°. The agreement is excellent. While these two papers do not completely settle the problem of calibration of the sintered glass diffusion membrane, they appear to represent a long step toward the solution. We have not attempted at this time, however, to recalculate the earlier data in the light of this new work. Hence the values given in the following sections must be considered in relation to the calibration on which they were based.

### The Diffusion Constants of the Amino Acids

It will be evident from equation 13 that the diffusion of an amino acid at a pH other than that of the isoelectric point will depend upon the nature of the other ions present in the solution. Since, in general, the mobilities of other ions will be greater than those of the relatively large ions of the amino acids, the diffusion of the salt of an amino acid will be more rapid than that of the isoelectric amino acid. This effect is illustrated by some experiments of McBain and Dawson<sup>32</sup>, where the rate of diffusion of glycine was increased about 10% by the addition of sufficient hydrochloric acid to change the pH from 6 to 4. Except in such cases as this, however, where acids or alkalies are added to the amino acid solution, the charge effects will be negligible for the monoamino-monocarboxylic acids. The isoelectric points lie so close to pH 6 that the concentrations of the ionic forms need not be considered. In this case, the diffusion coefficient will depend primarily upon the size and shape of the amino acid.

The diffusion coefficients of a number of the amino acids are given in Table 2. Although there is some disagreement as to the magnitude of the effect, all the amino acids studied so far (with the possible exception of arginine) show a decrease in the diffusion coefficient as the concentration increases<sup>6, 30</sup>. Consequently, the values given in Table 2 are those obtained by extrapolation to infinite dilution. This should make the results of free diffusion experiments comparable with those made by the disk method.

In several cases the diffusion measurements have been made at temperatures other than 20°, as indicated in the column headed "temperature". In order to compare results, all the values have been reduced to 20° with the aid of the relation

$$D\eta/T = \text{Constant} \quad (32)$$

obtained experimentally. It is a relation which follows from both equation 15 and from the theoretical treatment of Eyring<sup>33</sup>, and has been tested experimentally in a number of cases<sup>6, 34</sup>. In the case of the amino acids studied by Mehl and Schmidt<sup>6</sup> the following relations ( $D$  in  $\text{cm}^2/\text{sec}$ ) were found to hold:

$$\text{Glycine:} \quad D = 3.04 \times 10^{-10} (T/\eta)[1 - 0.0014(T - 274)]$$

$$dl\text{-Alanine:} \quad D = 2.57 \times 10^{-10} (T/\eta)[1 - 0.0014(T - 274)]$$

$$dl\text{-Valine:} \quad D = 2.17 \times 10^{-10} (T/\eta)$$

$$l\text{-Asparagine:} \quad D = 2.42 \times 10^{-10} (T/\eta)[1 - 0.0017(T - 283)]$$

$$dl\text{-Proline:} \quad D = 2.48 \times 10^{-10} (T/\eta)$$

Even in the case of asparagine, which showed the greatest deviation, the error involved in calculating a diffusion coefficient for  $30^\circ$  from measurements made at  $10^\circ$  with the help of equation 32 would only be about 2 per cent.

With the exception of that obtained by Ross and Salter, the values obtained for the diffusion coefficient of glycine are in reasonably good agreement if the Cohen and Bruins value for the diffusion coefficient of 0.1*N* potassium chloride is taken as the standard for calibrating the disk type diffusion cell. The deviation of the value of Ross and Salter for glycine is rather surprising in view of the excellent agreement between the values obtained by Salter for isoleucine and Polson for leucine. With respect to the other values, it will be seen that the agreement between the disk and free diffusion methods is again satisfactory if the same calibration is used. This would indicate that the Cohen and Bruins value for potassium chloride is satisfactory for the calibration of the sintered glass diffusion cells, but would leave discrepancies in certain other data unexplained.

In a general way, the diffusion coefficient and the molecular weight or molecular volume stand in the expected relation. The trend of decreasing diffusion coefficients with increasing molecular size is fairly regular, though there are some inversions in order. It is certainly not to be expected that equation 15 would be valid for molecules as small as the amino acids, and it may reasonably be expected that other factors such as the shape of the molecule will influence the diffusion coefficient. At present, however, there is no satisfactory method for treating this problem. (See however Kincaid, Eyring and Stearn<sup>33</sup>.)

### The Diffusion Constants of the Proteins

Because of the necessity for knowing the diffusion coefficient as well as



interpretation of these values for the diffusion of proteins doubly difficult. In the tabulation of the diffusion data for the proteins, account must again be taken of the method of calibration used.

In the case of the proteins, as was also indicated for the amino acids, diffusion potentials may have a considerable influence on the results. Making measurements at the isoelectric point of the protein will, of course, minimize this effect. In addition, because of the existence of an appreciable concentration of charged species in isoelectric solutions of proteins (Chapter 20), it is desirable to take further steps to reduce the diffusion potential. The expression for the diffusion potential in a mixture of salts is analogous to equation 13, with a term involving the reciprocal of the total salt concentration. Consequently, the diffusion potential may be reduced by increasing the total salt concentration without increasing the

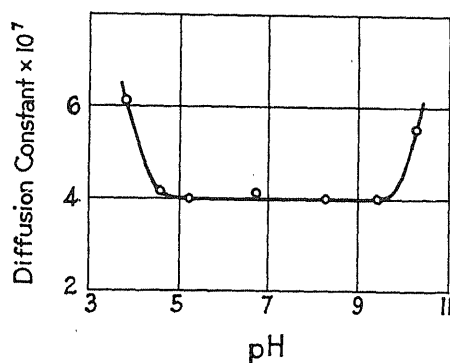


Figure 3. The dependence of the diffusion constant of phycoerythrin on pH. From Tiselius, A. and Gross, D., *Kolloid-Z.*, **66**, 11 (1934).

electrolyte concentration gradient. This is accomplished in practice by prolonged dialysis of the protein solution against the pure solvent, containing salt at the desired concentration. The effect of the diffusion potential at a pH removed from the isoelectric point is illustrated by the data of McBain, Dawson and Barker<sup>35</sup>. It is shown, for example, that the diffusion coefficient of egg albumin increases by about 50% when the pH is changed from 4.5 to 4.0. The effect of a neutral salt in reducing the diffusion potential and hence the diffusion coefficient is also shown by their data on egg albumin. The effect of neutral salt is also shown by the data of Lamm and Polson<sup>21</sup>. In the presence of a sufficient ionic strength, the diffusion coefficient of a protein may be constant over a wide pH range. This is shown by the data of Tiselius and Gross<sup>19</sup> on R-Phycoerythrin

The rise in  $D$  on the acid and alkaline sides of this range probably reflects a splitting of the molecule into smaller components.

In addition to the effect on the diffusion potential, added salt and buffer will alter the diffusion coefficient as a result of their effect on the viscosity of the medium through which diffusion is taking place. There may, as a matter of fact, be electroviscous effects, or large effects due to the aggregation or dissociation of protein molecules in salt solutions; but the former will probably be small, and the latter effects should be readily recognizable. Consequently, in order to compare diffusion measurements made in different kinds and concentrations of salts account is only taken of the effect of the salts on the macroscopic viscosity of the medium, and equation 32 is used to obtain a diffusion coefficient for pure water at 20°. The validity of equation 32 has been shown from measurements on R-Phycoerythrin by Tiselius and Gross<sup>19</sup> at 20° and 30°.  $D$  in water at 30° was found to be  $5.23 \times 10^{-7}$ . The value calculated for 30° by equation 32, from the observed value in water at 20° ( $4.00 \times 10^{-7}$ ), was  $5.18 \times 10^{-7}$ , which is well within the experimental error of the observed value. Polson<sup>36</sup> has also shown the validity of equation 32 by measurements on egg albumin at 15°, 20° and 25°.

It will be seen from Table 3 that some very striking concentration effects have been reported by Lamm and Polson<sup>21</sup> for egg albumin, serum albumin, and hemoglobin<sup>37</sup>. They point out the fact that the sedimentation constant has also been found to fall in dilute solutions of hemoglobin, indicating a dissociation of the molecule. Such observations on the two albumins are as yet lacking, but it seems possible that the same explanation may apply. At any rate, it does not seem desirable to extrapolate the data for the diffusion of the proteins to infinite dilution as was done for the amino acids. Where the diffusion has been measured over a range of concentrations, the value selected for comparison would preferably be at a concentration in the region where the diffusion coefficient is independent of the concentration. The same considerations would, of course, apply to the selection of values of the diffusion coefficient for use in conjunction with the sedimentation constant.

One point may be mentioned in connection with the values obtained by the disk method. An objection which has been most commonly made to the use of the porous disk in the study of proteins is that the disk may readily become clogged with precipitated protein, or that it may hinder the diffusion of the protein by wall effects. The results in Table 3 do not, however, show any tendency for the values obtained by this method to be lower than those obtained by free diffusion methods. The agreement

TABLE 3. DIFFUSION CONSTANTS OF PROTEINS WHICH HAVE BEEN STUDIED AT SEVERAL CONCENTRATIONS

Concentration (%)	Ref.	Solvent	Temp. (°C.)	$D \times 10^7$ (cm <sup>2</sup> /sec)	$D_{20} \times 10^7$ in H <sub>2</sub> O
CO-hemoglobin, horse					
1.00	(1)	0.01 <i>M</i> PO <sub>4</sub> pH 6.8	20	6.26	6.3 at 1%
0.80				6.27	
0.66				6.16	
0.50				9.68	
0.17				14.90	
1.00	(2)	0.040 <i>M</i> PO <sub>4</sub> pH 7.5	20		7.09
0.58					7.20
0.38					7.51
5.0	(3)	<i>M</i> /40 PO <sub>4</sub> 0.5% NaCl pH 6.8	0	4.07	8.5 at 2%
10.0				3.95	
20.0				3.61	
2.0	(4)	<i>M</i> /20 PO <sub>4</sub> pH 6.8	5	3.94	6.3 at 2%
CO-hemoglobin, human					
3.8	(5, 2)	0.0225 <i>M</i> PO <sub>4</sub> 0.1 <i>M</i> NaCl	20	6.73	6.9 at 1% to 4%
2.8				6.77	
0.82				6.77	
0.4				7.34	
0.2				7.34	
Egg albumin, chicken					
0.4	(1)	0.01 <i>M</i> Acetate 0.1 <i>M</i> NaCl pH 4.7	20	7.53	7.7 at 0.4%
1.4	(5)	0.04 <i>M</i> Acetate 0.2 <i>M</i> NaCl pH 4.6	20	7.44	8.4 at 0.5% 7.7 at 1%
0.91				7.50	
0.88				7.55	
0.83				7.52	
0.7				7.50	
0.5				8.20	
0.15				9.71	
1.00	(6)	H <sub>2</sub> O	25	11.3	
1.00	(6)	HCl to pH 4.5	25	11.2	9.8 at 1%
1.00	(6)	H <sub>2</sub> SO <sub>4</sub> to pH 4.8	25	11.2	
Serum albumin, horse*					
2.0	(5)	0.2 <i>M</i> Acetate 0.2 <i>M</i> NaCl pH 5.7	20	5.87	6.4 at 1%
0.93				5.95	
0.5				6.05	
Serum albumin, horse†					
0.850	(7)				6.0
0.425					6.1
Serum albumin, horse‡					
1.44	(7)				6.0
0.720					6.3

The measurements of Anson and Northrop (4) and of Zeile (3) were made employing the porous disk method. All of the other values given in this table were obtained by the method of free diffusion.

- (1) Tiselius, A., and Gross, D., *Kolloid. Z.*, 66, 11 (1934).
- (2) Polson, A., *Kolloid. Z.*, 87, 149 (1939).
- (3) Zeile, K., *Biochem. Z.*, 258, 347 (1932).
- (4) Anson, M. L., and Northrop, J. H., *J. Gen. Physiol.*, 20, 575 (1937).
- (5) Lamm, O., and Polson, A., *Biochem. J.*, 30, 528 (1936).
- (6) McBain, J. W., Dawson, C. R., and Barker, H. A., *J. Am. Chem. Soc.*, 56, 1021 (1934).
- (7) Kekwick, R. A., *Biochem. J.*, 32, 552 (1938).
- (8) Svedberg, T., and Eriksson-Quensel, I.-B., *Tab. Biol.*, 11, 351 (1935/36).
- (9) Polson, A., *Nature*, 127, 740 (1931).



TABLE 4. DIFFUSION CONSTANTS OF CERTAIN PROTEINS, CORRECTED TO STANDARD CONDITIONS (WATER AT 20°)

Protein	Ref.	Conc. (%)	pH	Temp. (°C.)	$D_{20} \times 10^7$ in H <sub>2</sub> O
Amandin, Almond	(9)			20	3.6
Bence-Jones $\beta$	(2)	1.0		20	7.3
Canavalin	(10)	0.68			5.1
Catalase	(11)	1			4.1
Concanavalin A	(10)	0.7			5.6
Concanavalin B	(10)	0.4			7.4
Cytochrome-C	(2)	0.5		20	10.1
Edestin, Hemp	(8)	0.88		20	4.0
Egg Albumin, Hen	(5)	1	4.6	20	7.7
Erythrocrucorin, Lampetra	(2)			20	10.65
Erythrocrucorin, Planorbis corneus	(5, 2)	0.44	6.7	20	1.96
Erythrocrucorin, Lumbricus	(2)	0.5		20	1.81
Excelsin, Brazil nut	(8)			20	4.3
Gliadin I	(5)	0.616		20	4.96
Gliadin II	(5)	0.62		20	5.85
Gliadin III	(5)			20	6.72
Gliadin A	(12)			25	4.46
Gliadin B	(12)			25	5.62
Gliadin C	(12)			25	6.03
Hemocyanin, Panularis vulgaris	(2)				
	(14)		6.8	20	3.4
Hemocyanin, Helix pomatia	(2)				
	(14)	0.8	9.7	20	2.23
			8.6		2.06
			6.8		1.38
Hemocyanin, Busycon caliculatum	(2)				
	(14)		9.6	20	3.29
			6.8	main comp.	1.38
			6.8	minor comp.	1.17
Hemocyanin, Eledone moschata	(2)				
	(14)	1.0	11.6	20	2.16
			6.8		1.64
Hemocyanin, Nephrops norvegicus	(2)				
	(14)	0.8	6.8	20	2.79
Hemocyanin, Homarus vulgaris	(2)				
	(14)	0.67-1.39	6.8	20	2.78
Hemocyanin, Helix nemoralis, dissn. comp.	(2)				
	(14)	0.8	8.6	20	1.92
Hemocyanin, Octopus vulgaris	(2)				
	(14)	0.8	6.8	20	1.65
Hemocyanin, Rossia owenii	(2)				
	(14)		6.8	20	1.58
CO-Hemoglobin, Horse	(1, 4)	1	6.8	5, 20	6.3
in 4M urea	(13)			20	7.81
in 6.46M acetamide	(13)			20	7.68
in 4.55M formamide	(13)			20	7.81
CO-Hemoglobin, Human	(5)	1-4	6.8	20	6.9
CO-Hemoglobin, Lampetra	(2)	0.32	6.7	20	10.65
Insulin	(2)	0.2		20	8.2
Lactalbumin $\alpha$	(2)			20	10.6
Lactoglobulin	(5, 2)	1	6.9	20	7.2
Myoglobin	(2)	0.4	7.1	20	11.25
Pepsin	(2)	1	2.0-5.4	20	9.0
Phycocyan, Ceramium	(2)	0.2	5.0	20	4.15
dissn. comp.					
Phycocyan, R.	(1)	0.15	5.0	20	4.05
Phycocerythrin, Ceramium	(2)	0.4	6.67	20	4.06

between the two methods is, on the whole, reasonably good. The higher value of McBain, Dawson and Barker<sup>35</sup> for egg albumin, which cannot be brought into agreement with the value from free diffusion, even allowing for differences in the method of calibration, may possibly be due to the fact that salt was not used to reduce the diffusion potential in the measurements at the isoelectric point.

Neurath and Saum<sup>38</sup> have studied the diffusion constant of serum albumin in urea solutions of varying concentration, and have found a very

TABLE 5. DIFFUSION CONSTANTS OF CERTAIN PROTEINS CORRECTED TO STANDARD CONDITIONS (WATER AT 25°)

Protein	Ref.	Conc. (%)	pH	$D_{25} \times 10^7$ in H <sub>2</sub> O
Tobacco Mosaic Virus	(15)	0.2-0.5	6.8	0.3
Tomato Bushy Stunt Virus	(16)	0.2-0.4	7.1	1.31
Pepsin (A)	(17)	1.0	5.0	9.17
Pepsin (B)	(17)	0.7	5.0	9.96
Serum Albumin (Hewitt's method)	(17)	0.7	5.0	6.95
Serum Albumin (McMeekin's method)	(17)	0.7-1.0	5.0	7.36
Serum Albumin* (seroglycoid)	(17)	0.5	5.0	6.93
Pseudoglobulin G I	(17)	0.5-1.5	5.5	4.61
Pseudoglobulin G II	(17)	0.5-1.5	5.5	4.75
Pseudoglobulin G III	(17)	0.5-1.5	5.5	4.55
Ribonuclease†	(18)	1.0		13.6

\* Carbohydrate content 4.4 mg per gm of protein.

† Value is for 0.5M (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> at 25°; not corrected to water at 25°.

See notes to table 3.

TABLE 6. DIFFUSION CONSTANTS AND SHAPE FACTORS OF SERUM ALBUMIN IN UREA SOLUTIONS

(Molecular weight 67,100;  $D_0^{25} = 8.22 \times 10^{-7}$ ;  $a$  = length of short axis of molecules)

Urea (M)	$D \times 10^7$	$\frac{D_0}{D}$	$\frac{a}{b}$
0	6.85	1.20	4.3
0.5	6.20	1.33	6.1
1.5	6.08	1.35	6.5
3.0	5.69	1.44	8.0
4.5	4.45	1.85	16.5
6.0	4.27	1.93	18.3
6.66	4.15	1.98	19.4

Values of  $a/b$  are calculated from equation 27, assuming no hydration.  
From Neurath, H. and Saum, A. M., *J. Biol. Chem.*, 128, 359 (1939).

marked decline in  $D$  with increasing urea concentrations, even when the observed value is corrected for changing viscosity of the medium by equation 32.

Since the osmotic pressure measurements of Burk<sup>39</sup> show no change in the molecular weight in urea solutions of the same concentration, the fall in value of  $D$  must presumably be ascribed to an increase in the asymmetry of the molecule, which may be evaluated by means of equation 27 (Table 6). In contrast to urea solutions serum albumin shows no signifi-

cant change in the corrected diffusion constant when dissolved in sodium salicylate, even up to 3.5 molar, although the correction factor for viscosity of the medium in the latter case amounts to 700 per cent. The increasing specific viscosity (Chapter 21) of serum albumin, in solutions of increasing urea concentration, parallels the decreasing diffusion constant, and also points clearly to increasing molecular asymmetry.

Since recent work indicates<sup>40, 41</sup> that serum albumin as ordinarily prepared does not consist of a single molecular species, the quantitative interpretation of the data of Neurath and Saum may require some future revision. The general character of their results, however, is unmistakable<sup>42</sup>

The important work of Steinhardt on the diffusion and sedimentation of hemoglobin and pepsin in solutions of urea and other amides is discussed in Chapter 19.

<sup>40</sup> Hewitt, L. F., *Biochem. J.*, **30**, 2229 (1936).

<sup>41</sup> McMeekin, T. L., *J. Am. Chem. Soc.*, **61**, 2884 (1939); **62**, 3393 (1940).

<sup>42</sup> Recently Neurath and his collaborators (24) have studied some of the more highly purified serum albumin fractions (Table 5), but as yet they have reported results only for aqueous solutions.

# Chapter 19

## Sedimentation and Diffusion in Centrifugal Fields: Molecular Weights of Proteins

BY JOHN T. EDSALL

### Effect of a Centrifugal Field in Modifying the Diffusion Process

In certain systems of great importance for the study of protein molecules, the process of free diffusion is modified by the presence of an external force which produces a directed flow of the solute molecules<sup>1</sup>. If this force (acting in the  $x$  direction) is not too large, it tends to make the solute molecules move parallel to the  $x$  axis with constant velocity; the external force,  $\phi$ , per gram mole of solute being balanced by the frictional resistance of the medium. If the resulting velocity is denoted by  $dx/dt$ , then the frictional resistance is  $F \frac{dx}{dt}$ , and

$$\frac{dx}{dt} = \frac{\phi}{F} = \frac{\phi D}{RT} \quad (1)$$

where  $F$  is the molar frictional coefficient of the solute, which, as has already been stated (Chapter 18, equations 11 and 24) is inversely proportional to its diffusion constant. The flow of solute across an area  $A$ , due to the external force, at a point where the concentration is  $C$ , is  $\frac{dx}{dt} CA$  per unit time. On the other hand, the flow due to free diffusion (Chapter 18, equation 1) is  $-DA \frac{dc}{dx}$ . Therefore the net rate of flow across the area is

$$\begin{aligned} \frac{dS}{dt} &= A \left[ C \frac{dx}{dt} - D \frac{dc}{dx} \right] \\ &= A \left[ \frac{\phi C}{f} - D \frac{dc}{dx} \right] \\ &= DA \left[ \frac{\phi C}{RT} - \frac{dc}{dx} \right] \end{aligned} \quad (2)$$

which is a generalized form of Fick's first law.

<sup>1</sup> The theoretical discussion in this chapter is given in relatively simple form, in order to emphasize the main points. A more complete and critical discussion is given by Syedberg,

The rate of change of concentration with time at any point is found from this equation to be

$$\frac{dc}{dt} = -\frac{1}{f} \frac{\partial(\phi c)}{\partial x} + D \frac{\partial^2 c}{\partial x^2} = D \left[ -\frac{1}{RT} \frac{\partial(\phi c)}{\partial x} + \frac{\partial^2 c}{\partial x^2} \right] \quad (3)$$

which may be derived from equation 32 of Chapter 18 by the same process involved in deriving Fick's second law (Chapter 18, equation 3) from his first law (Chapter 18, equation 1)<sup>2</sup>. It should be noted that equations 2 and 3 presuppose that  $D$  is independent of the concentration,  $C$ , of the solute over the range studied, and must be modified if this assumption is invalid.

### Sedimentation Equilibrium

Equations 2 and 3 have been found to be of importance in certain electrophoresis experiments, in considering the changes produced by free diffusion in the migration of proteins in an electric field. Their principal application hitherto, however, has been in the use of the ultracentrifuge, where the external force  $\phi$  is centrifugal, and the results of the experiment may be used to obtain the molecular weight of the protein. Two methods have been employed in ultracentrifugal studies for attaining this end. In the first, the method of sedimentation equilibrium, the centrifugal force is relatively low, and the experiment is prolonged until the distribution of protein throughout the column of liquid in the ultracentrifuge has reached a steady state; that is, the net flow across any cross-section perpendicular to the  $x$  axis is zero. Then, in equation 2,  $ds/dt = 0$  at all points, and

$$\frac{\phi c}{RT} = \frac{dc}{dx} \quad (4)$$

The centrifugal force,  $\phi$ , per gram mole, is equal to the outward acceleration produced by the rotation of the centrifuge, multiplied by the difference between the mass of one mole of protein and the mass of the solvent displaced by it. The acceleration is  $\omega^2 x$ , where  $\omega$  is the angular velocity of the centrifuge and  $x$  is the distance from the center of rotation, and

$$\phi = \omega^2 x M(1 - V\rho) \quad (5)$$

where  $M$  is the molecular weight of the protein,  $V$  its partial specific volume—that is, the increment in volume when one gram of dry protein is added to a large amount of solvent—and  $\rho$  is the density of the solution. Substituting 5 in 4 and rearranging terms

$$\frac{dc}{c} = \frac{\omega^2 M(1 - V\rho)}{RT} x dx \quad (6)$$

which gives on integrating between two points distant  $x_1$  and  $x_2$  cm from the axis of rotation:

$$M = \frac{2RT \ln c_2/c_1}{\omega^2(1 - V\rho)(x_2^2 - x_1^2)} \quad (7)$$

Thus, if a state of equilibrium has been reached and if the solute is monodisperse, its molecular weight may be evaluated by determining the relative concentrations at two distances from the center of rotation. If a series of values of  $c$  as a function of  $x$  is obtained, any inhomogeneity with respect to molecular weight will be shown by a drift of the calculated values of  $M$  with increasing values of  $x$ . The absence of drift is therefore a reliable indication that the solute is uniform with respect to molecular weight. It does not, of course, establish the chemical homogeneity of the solute in other respects; indeed mixtures of naturally occurring proteins may be obtained, which are uniform with respect to molecular weight, but are separable by electrophoresis or by chemical methods. The serum albumins and serum globulins furnish illustrations of such systems.

### Sedimentation Velocity

If centrifugal fields of much higher intensity are employed, the process of sedimentation is much more rapid than that of free diffusion; the protein molecules, being denser than the solvent, move outward, away from the center of rotation, and a fairly sharp boundary between solvent and protein solution is formed. Observation of the rate of motion of this boundary is the basis of the sedimentation velocity method.

During the course of the run, the sedimenting boundary becomes gradually blurred by diffusion of solute back into the pure solvent; the method of correction for this effect, and the calculation of the true position for the sedimenting boundary, have been described by Svedberg and Pedersen (reference 1, p. 18).

It is important in practice, so that motion of the molecules may follow radial lines, that sedimentation should take place in a sector-shaped cell. Therefore the area across which sedimentation and diffusion takes place increases with increasing distance from the center of rotation. This introduces additional terms into equations 2 and 3, which have been fully discussed by Svedberg and Pedersen; the final equations, 7 and 10, from which the molecular weight is calculated, are unaffected.

At the beginning of an experiment, the protein molecules, starting from

resistance of the medium, as described in equation 1. Then the sedimentation velocity, from 1, is:

$$\frac{dx}{dt} = \frac{\phi D}{RT} = \frac{\omega^2 x M(1 - V\rho)D}{RT} \quad (8)$$

The quantity  $s$ :

$$s = \frac{dx/dt}{\omega^2 x} \quad (9)$$

is known as the sedimentation constant; it is the sedimentation velocity for unit centrifugal acceleration, and has the dimensions of time. For the proteins hitherto studied it lies between 1 and  $200 \times 10^{-13}$  sec. It has been recently recommended\* that a sedimentation constant of  $10^{-13}$  sec be denoted as one Svedberg unit ( $S$ ), and that all sedimentation constants be reported in Svedberg units.

On substituting 9 in 8, and rearranging, we obtain for the molecular weight:

$$M = \frac{RTs}{D(1 - V\rho)} \quad (10)$$

Thus, for the determination of molecular weight by the sedimentation velocity method, independent studies of sedimentation and diffusion on the same solute must be carried out. The values for  $s$  and  $D$  as reported from Svedberg's laboratory are generally given for the protein dissolved in water at  $20^\circ$ . If the actual determinations are made at a different temperature and in a different solvent, the values obtained are corrected to this standard state. The corrected values for  $D$  are obtained by the use of equation 32, Chapter 18; those for  $s$  from the formula

$$s_{20} = s \frac{\eta}{\eta_{20}} \frac{(1 - V_{20}\rho_{20})}{(1 - V\rho)} \quad (9a)$$

where  $s$  is the observed sedimentation constant,  $\rho$  and  $\eta$  are the density and viscosity of the solvent at the temperature of the experiment, and  $V$  is the partial specific volume of the protein under the conditions of the experiment;  $\eta_{20}$ ,  $\rho_{20}$ , and  $V_{20}$  are the corresponding quantities for water at  $20^\circ$ .

It is important, especially if the molecules studied are far from spherical in shape, to determine both  $s$  and  $D$  over a considerable range of concentrations. It has been found by Signer and Gross<sup>4</sup> that the sedimentation

\* At a symposium on the ultracentrifuge held by the New York Academy of Sciences, November, 1941. See MacInnes, D. A. *Annals N. Y. Acad. Sci.* 43, 176 (1942).

constants of polystyrenes of high molecular weight, in organic solvents, vary greatly with concentration, the value of  $s$  increasing with the dilution;  $1/s$  is approximately a linear function of the concentration.\* Behavior at least qualitatively similar is characteristic of all natural and synthetic linear high polymers<sup>5</sup>. Recently calf thymus nucleohistone<sup>6</sup> has been found to show similar behavior;  $s_{20}$ , expressed in Svedberg units, varies from 12.5 at 0.82% protein to 24.7 at 0.10% protein. The extrapolated value at infinite dilution is 31. The frictional ratio is 2.5, indicating the extreme asymmetry of the molecules. Nevertheless, sedimentation velocity and diffusion experiments gave results in good agreement with sedimentation equilibrium studies, the molecular weight being between 2,000,000 and 2,300,000.

Equation 7, for sedimentation equilibrium, has also been derived by purely thermodynamic reasoning (reference 1, p. 8 and p. 48 ff.); the molecular weight so determined is equivalent to one determined by measurement of the colligative properties of the solution. The sedimentation velocity method involves the additional assumption of the validity of Fick's laws of diffusion for the solute molecules<sup>7</sup>. The theoretical foundation of this method is thus slightly less rigorous than that of the equilibrium method. The former has, however, two very great practical advantages: (1) sedimentation velocity experiments are relatively rapid, requiring generally only a few hours, while the attainment of sedimentation equilibrium may take several days or even weeks. The shorter time involved in the former type of experiment is not only a convenience, but it also minimizes the danger that the protein molecules may decompose or be otherwise altered during the experiment; (2) in a polydisperse system, two or more separate boundaries are set up during a sedimentation velocity experiment, corresponding to the sedimentation of components of differing size or shape. Thus the relative amounts and sedimentation constants of the different components may be determined, provided the components differ sufficiently in size or shape. The ability to separate different components depends upon the resolving power of the centrifuge, which is proportional to  $\omega^2 \times h$ , where  $h$  is the length of the column of liquid in which sedimentation occurs. In the latest and most powerful centrifuge designed by Svedberg, it is possible to separate two molecular species with molecular weights of 20,000 and 40,000 respectively<sup>8</sup>.

Unfortunately, however, knowledge of the sedimentation constant alone is not sufficient to determine the molecular weight. It is apparent from

\* This linear relation between  $1/s$  and  $c$  is purely empirical and cannot be expected to hold in all cases.

<sup>5</sup> Kraemer, E. O., and Nichols, J. B., Ref. 1, p. 416; Signer, R., Ref. 1, p. 431.



equation 10 that the diffusion constant is of equal importance; and diffusion constants of proteins are in general much less accurately known than sedimentation constants. Determination of a diffusion constant requires a prolonged experiment, lasting several days. The danger of alteration of the protein during the experiment is therefore considerable, as it is during a sedimentation equilibrium determination. For this and perhaps for other technical reasons, the probable errors in the  $D$  values to be employed in equation 10 largely determine the probable errors in the estimated molecular weights.

### Frictional Ratios and Their Relation to Asymmetry and Hydration

Svedberg has also shown how a factor which is a function of molecular asymmetry may be derived from a combination of sedimentation velocity (or diffusion) and sedimentation equilibrium data. If  $M_s$  is the molecular weight from sedimentation equilibrium<sup>9</sup>, then the molecular frictional coefficient from sedimentation data is

$$f_s = \frac{M_s(1 - V\rho)}{Ns} \quad (11)$$

or from the diffusion constant (Chapter 18, equation 24a)

$$f_d = \frac{kT}{D} \quad (12)$$

It has been found experimentally that  $f_s = f_d$ . The molar frictional coefficient for a sphere of the same molecular weight may be calculated from equation 15 of Chapter 18, and the formula for the volume of a sphere.

$$f_0 = 6\pi\eta rN = 6\pi\eta N \left( \frac{3MV}{4\pi N} \right)^{1/3} \quad (13)$$

where  $V$  is again the specific volume of the protein. Then

$$\frac{f}{f_0} = \frac{f_s}{f_0} = \frac{f_d}{f_0} = \frac{D_0}{D} \quad (14)$$

Numerical values of  $f/f_0$  have been obtained in Svedberg's laboratory for many proteins, and have been found in practice always to be greater than unity. Values greater than unity may be explained as due either to asymmetry or to hydration of the molecule, or more probably as due to a combination of these two effects. Thus, following Oncley<sup>10</sup>, we may write the observed  $f/f_0$  value as the product of two factors:

$$\frac{f}{f_0} = \left( \frac{f}{f_0} \right) \left( \frac{f_e}{f_0} \right) \quad (15)$$

where the first denotes the influence of hydration, and the second the influence of asymmetry. If the form of a protein molecule may be approximated by that of an ellipsoid of revolution, we may employ the equation of Perrin\* to obtain  $f_e/f_0$  as a function of the axial ratio  $a/b$ . The hydration factor,  $f/f_e$  is given by the equation<sup>10a</sup>

$$\frac{f}{f_e} = \left(1 + \frac{w}{V\rho}\right)^{1/3} \quad (16)$$

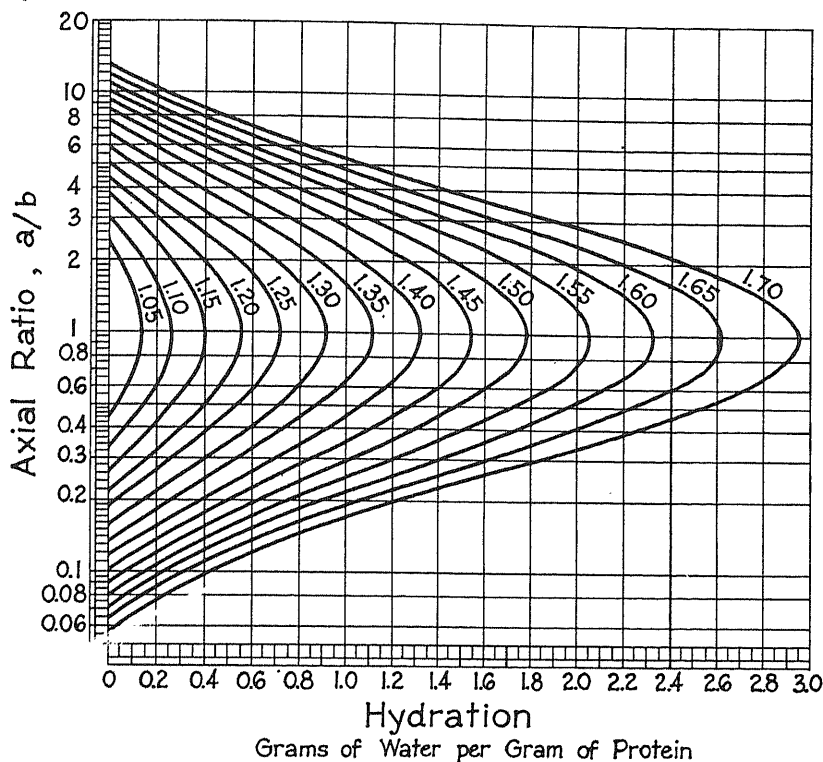


Figure 1. Contour lines denote  $f/f_0$  values.

where  $w$  is the hydration, expressed as grams of water bound by one gram of protein. Oncley<sup>10</sup> has given a figure (Fig. 1) in which the hydration is plotted against the axial ratio, on a logarithmic scale, the values of the frictional ratio being given by contour lines. From this diagram it is easy to read off all the values of hydration and asymmetry compatible with a

### Experimental Methods

The sedimentation equilibrium of colloidal particles in a gravitational field played a part of fundamental importance in colloid chemistry a generation ago<sup>11</sup>. Much larger forces were required, however, to produce a readily measurable sedimentation of particles of the size of protein molecules; and this vast field was first opened to investigation by the great work of Svedberg in the development of the ultracentrifuge. The details of construction and operation of the ultracentrifuge have been thoroughly covered by Svedberg and Pedersen<sup>1</sup>, and will not be described here. Svedberg has constructed two types of instrument; one for speeds which will produce centrifugal fields from 50 to 18,000 times the force of gravity—this type is particularly adapted for sedimentation equilibrium determinations<sup>12</sup>; the other for centrifugal forces from 8000 to 500,000 times the force of gravity, or in some cases even higher. Some of the latest developments in the use of these ultracentrifuges have recently been described from Svedberg's laboratory<sup>1, 13</sup>.

Svedberg's high-speed ultracentrifuge is driven by an oil turbine, and the motor runs in an atmosphere of hydrogen at a pressure of a few millimeters of mercury. The presence of the hydrogen is essential to conduct away the heat which develops in the rotor during the run. The air-driven type of ultracentrifuge is based on somewhat different mechanical principles, and has been developed from the spinning top of Henriot and Huguenard<sup>14</sup>. During the last few years this type of instrument has been perfected especially by Beams and Pickels<sup>15</sup>, Bauer and Pickels<sup>16</sup>, and Wyckoff and Lagsden<sup>17</sup>. A blast of air turns the upper portion of the instrument, from which the rotor is suspended by a piece of piano wire. The piano wire is completely surrounded by packing which fits very closely and just leaves it free to turn, the very small gap between wire and casing being sealed by oil. The rotor is completely encased and runs in a high vacuum; heating effects are extremely small. This type of ultracentrifuge is useful not only for molecular weight determinations, but for the concentration of viruses and other very large protein molecules in preparative work. The maximum centrifugal force and resolving power attainable with this type of instrument are at present somewhat less than with Svedberg's ultra-

<sup>11</sup> The work of J. Perrin, Svedberg and others in this field is summarized by T. Svedberg in his "Colloid Chemistry," development of the detail the early phases of the

<sup>12</sup> For very large sedimentation velocity studies; *stunt virus protein, Biochem. J.*

<sup>13</sup> Boestau, G., Pedersen, K. O., and Svedberg, T., *Rev. Sci. Instruments*, 9, 346 (1938); Tiselius, A., Pedersen, K. O., and Svedberg, T., *Nature*, 140, 848 (1937).

<sup>14</sup> Henriot, E., and Huguenard, E., *Compt. rend. Acad. Sci.* 180, 1260 (1905).

centrifuge; but for all except the smallest protein molecules it is very effective, and it has the advantage of being rather less expensive to construct. Recent improvements in this instrument have been described by Beams, Linke and Sommer<sup>18</sup>; and Beams<sup>19</sup> has described a "tubular" vacuum type centrifuge for continuous flow, which may prove very useful in preparative work. Other recent technical improvements have been described by Hughes, Pickels and Horsfall<sup>20</sup> and Pickels<sup>21</sup>.

Recently Beams<sup>22</sup> has developed an electrically driven, magnetically supported, vacuum type ultracentrifuge. The system of controls employed maintains the speed of rotation constant within 0.05%, and it requires far less attention from the operator than does the air-driven ultracentrifuge. He also reports, in the same communication, a most interesting new development in his laboratory, in which the rotor is attached to a vertical iron rod, which is suspended by the coaxial field of a solenoid. The current through the solenoid is so regulated as to maintain the rotor at a constant height, free from mechanical contact with anything. The rotor is thus spun in a vacuum by a rotating magnetic field. McHattie, in Beams's laboratory, has spun a 3/32" steel ball to 6,600,000 rpm, giving a centrifugal force of 58 million times gravity. The future developments of this method may be very great.

Another type of spinning top ultracentrifuge has been developed by McBain and his collaborators<sup>23</sup>. The instrument is very simple and inexpensive to construct, and may lead to important developments, but the mechanical difficulties involved in its operation have not yet been completely resolved, and it has not yet provided values for the molecular weights of proteins comparable in reliability to those given by the other types of ultracentrifuge.

Several different optical methods have been employed in observing sedimentation velocity: the light absorption method, using light of a wave length absorbed by the protein; the refractometric method of Lamm; and the "schlieren" method. All these are fully discussed by Svedberg and Pedersen (1).

### Results of Ultracentrifugal Studies on Proteins

The results at present available on the molecular weights of proteins as derived from sedimentation and diffusion data are given in Table 1. These data, and the original communications from which they are derived, have supplied material leading to a number of general conclusions important for protein chemistry.

<sup>18</sup> Beams, J. W., Linke, F. W., and Sommer, P., *Rev. Sci. Instruments*, **9**, 248 (1938).

TABLE 1. MOLECULAR WEIGHTS AND FRICTIONAL RATIOS OF PROTEINS  
(This table, except for some modifications and additions, is taken from Svedberg and Pedersen ("The Ultracentrifuge" p. 406)

Protein	$V_{20}$	$s_{20}$	$D_{20}$	$M_s$	$M_e$	$f/f_0^*$	References to:				
							$V_{20}$	$s_{20}$	$D_{20}$	$M_s$	$M_e$
Erythrocyruorin (Lampetra).....	0.751	1.87	10.7	17,100	19,000	1.16	(1)	(1)	(2)	(3)	(1)
Lactalbumin.....	(0.751)	1.9	10.6	17,400		1.16		(4)	(2)	(3)	
Cytochrome c.....	0.707	1.9	10.1	15,600		1.29	(5)	(6)	(2)	(7)	
Myoglobin.....	0.741	2.04	11.3	16,900	17,500	1.11	(8)	(2)	(2)	(7)	(2)
Bacillus Phlei protein.....	0.748	1.8	10.2	17,000		1.22	(9)	(9)	(9)	(9)	
Ribonuclease.....	0.709 <sup>a</sup>	1.85 <sup>a</sup>	13.6 <sup>a</sup>	12,700	13,000	1.04	(10)	(10)	(10)	(10)	(10)
Gliadin**.....	0.722 <sup>b</sup>	2.1	6.72	27,500 <sup>b</sup>	27,000	1.60	(11)	(11)	(12)	(12)	(11)
Hordein.....	(0.729)	2.0	6.5	27,500		1.64		(13)	(13)	(13)	
Zein.....	(0.71)	1.9	4.0	40,000		2.4		(14)	(14)	(14)	
Erythrocyruorin (Arca).....	(0.751)	3.46			33,500	(0.99)		(15)			(1)
Erythrocyruorin (Chironomus).....	(0.751)	2.0			31,500	1.63		(15)			(1)
Lactoglobulin.....	0.7514	3.12	7.3	41,500	38,000	1.26	(16)	(16)	(2)	(16)	(16)
Pepsin.....	(0.750)	3.3	9.0	35,500	39,000	1.08		(17)	(2)	(18)	(19)
Insulin.....	0.749	3.5	8.2	41,000	35,000	1.13	(20)	(20)	(2)	(3)	(20)
Bence-Jones $\alpha$ .....	0.749	3.7 <sup>c</sup>			35,000	(0.96)	(21)	(21)			(21)
Bence-Jones $\beta$ .....	(0.749)	2.8	7.3	37,000		1.31		(6)	(2)		
Ovalbumin.....	0.749	3.55 <sup>d</sup>	7.8	44,000	40,500	1.16	(22) <sup>b</sup>		(12)	(12)	(23)
Human tuberculosis bacillus protein.....	0.70	3.3	8.2	32,000		1.25	(9)	(9)	(9)	(9)	
Concanavalin B.....	0.73	3.5	7.4	42,000		1.25	(24)	(24)	(24)	(24)	
Crotoxin.....	0.704	3.14	8.6	30,000	30,500	1.22	(25)	(25)	(25)	(25)	(25)
Oxytocic pressor hormone.....				31,000		1.18				(26)	
Hemoglobin (horse).....	0.749	4.41	6.3	68,000	68,000 <sup>e</sup>	1.24	(27)	(6)	(28)		(27)
Hemoglobin (man).....	(0.749)	4.48	6.9	63,000		1.16		(23)	(12)	(12)	
Serum albumin (horse).....	0.748	4.46 <sup>f</sup>	6.1 <sup>f</sup>	70,000 <sup>f</sup>	68,000	1.27	(29)	(30)	(30)	(30)	(29)
Yellow enzyme.....	0.731	5.76	6.3	82,000	78,000	1.17	(31)	(31)	(31)	(31)	(31)
Diphtheria toxin.....	0.736	4.6	6.0	74,000		1.22	(32)	(32)	(32)	(32)	
Concanavalin A.....	0.73	6.0	5.6	96,000		1.26	(24)	(24)	(24)	(24)	
Canavalin.....	0.73	6.4	5.1	113,000		1.31	(24)	(24)	(24)	(24)	
Diphtheria antitoxic pseudoglobulin (pepsin-treated).....	0.745	5.7	5.8	98,000		1.14	(32)	(32)	(32)	(32)	
Diphtheria antitoxin.....	0.749	5.5	5.76	90,500		1.23	(33, 34)	(33, 34)	(33, 34)	(33, 34)	
Serum globulin (horse) <sup>g</sup> .....	0.745	7.1	4.05	167,000	150,000	1.44	(29)	(35)	(2)	(3)	(19)
Serum globulin (man) <sup>h</sup> .....	(0.745)	7.1	3.84	176,000		1.49		(36)	(36)	(36)	
Antipneumococcus serum globulin (rabbit).....	(0.745)	6.5 <sup>i</sup>	3.9 <sup>i</sup>	158,000 <sup>i</sup>		1.52		(36)	(36)		



TABLE 1—Continued

Protein	$V_{20}$	$s_{20}$	$D_{20}$	$M_s$	$M_e$	$f/f_0^*$	References to:				
							$V_{20}$	$s_{20}$	$D_{20}$	$M_s$	$M_e$
Antipneumococcus serum globulin (pig)	(0.715)	18.0	1.64	930,000		2.02		(36, 48)	(36, 48)	(36, 48)	
Erythrocrucorin (Planorbis).....	0.745	33.7	1.96	1,630,000	1,540,000	1.39	(1)	(1, 15)	(12)	(12)	(19)
Hemocyanin (Calocaris).....	(0.740)	34			1,330,000	1.22		(15)			(45)
Hemocyanin (Octopus).....	0.740	49.3	1.65	2,800,000		1.38	(49)	(45)	(41)	(41)	
Hemocyanin (Eledone).....	(0.740)	49.1	1.64	2,800,000		1.39		(45)	(2)	(18)	
Nucleohistone (calf thymus).....	0.658	31.0	0.93	2,300,000	2,000,000	2.5	(50)	(50)	(50)	(50)	(50)
Erythrocrucorin (Arenicola).....	(0.740)	57.4			3,000,000	1.25		(51)			(19)
Chlorocrucorin (Spirographis).....		55.2						(19)			
Hemocyanin (Rossia).....	(0.740)	56.2	1.58	3,300,000		1.36			(2)	(18)	
Erythrocrucorin (Lumbrius).....	0.740	60.9	1.81	3,150,000	2,950,000	1.21	(51)	(51)	(2)	(3)	(19)
Hemocyanin (Helix pomatia, main component)	0.738	68.9	1.38	6,600,000	6,700,000	1.24	(52)	(45)	(2)	(18)	(45)
Hemocyanin (Busycon, main component).....	(0.738)	101.7	1.38	6,800,000		1.23		(45)	(45)	(45)	
Bushy-stunt virus....	0.739	146			7,600,000	1.09	(53)	(53)			(53)
Bushy-stunt virus....		132	1.15	10,600,000		1.27		(54, 55)	(58)	(58)	
Tobacco mosaic virus.		103						(54)			
Hemocyanin (Busycon, aggregation component).....	(0.738)	130	1.17	10,000,000				(45)	(45)	(45)	
Rabbit papilloma virus.....	0.756	280	0.585	47,100,000		1.49	(56)	(56)	(50)	(56)	

$V_{20}$  = partial specific volume of the protein. (The values in brackets are assumed values based on experimental determinations on closely related proteins.)

$s_{20}$  = sedimentation constant in Svedberg units reduced to water at 20°.

$D_{20}$  = diffusion constant in units of  $10^{-7}$  reduced to water at 20°.

$M_s$  = molecular weight determined by sedimentation measurements.

$M_e$  = molecular weight determined by other measurements.

$f/f_0$  = ratio of frictional constant of the protein to that of an unsolvated spherical particle of the same mass.

Values for  $M_s$  and  $M_e$  beginning with 1 are given to three significant figures; those beginning with 2 to 4 are rounded off to 0 or 5 in the third figure; and those beginning with 5 to 9 are given to two significant figures.

\* The reported  $f/f_0$  values are probably significant only within  $\pm 0.03$  and are sometimes even less accurate than this.

\*\* McCalla and Gralén<sup>57</sup> have studied wheat gluten in sodium salicylate solutions. The preparations were polydisperse; the weight-average molecular weight of the most soluble fraction was 44,000, but the minimum molecular weight of particles in this fraction may have been as low as 35,000.

† Theorell and Åkesson (Theorell, H. and Åkesson, Å., *J. Am. Chem. Soc.*, 63, 1804 (1941)) have obtained

TABLE 1—Continued

- <sup>c</sup> Recalculated to water basis.  
<sup>d</sup> Mean value obtained by a number of different observers in Upsala.  
<sup>e</sup> Mean of two determinations on CO-hemoglobin and two on methemoglobin.  
<sup>f</sup> Mean values for the A and B fractions prepared by Kekwick (1938).  
<sup>g</sup> Total serum globulin obtained by half saturation with ammonium sulfate.  
<sup>h</sup> Human serum.
- (1) (1930).
  - (2) Polson, A. G., *Proc. Roy. Soc. Med.*, 56, 1700 (1934).
  - (3) Svedberg, T., *Naturwissenschaften* (1937).
  - (4) Kekwick, R. A., Unpublished.
  - (5) Theorell, H., *Biochem. Z.*, 285, 207 (1936).
  - (6) Pedersen, K. O., and Andersson, K. J. I., Unpublished.
  - (7) Svedberg, T., *Chem. Rev.*, 20, 81 (1937).
  - (8) Theorell, H., *Biochem. Z.*, 268, 46 (1934).
  - (9) Pedersen, K. O., and Tiselius, A., *J. Exptl. Med.*, 68, 413 (1938); *Am. Rev. Tuberc.*, 38, 399 (1938).
  - (10) Rothen, A., *J. Gen. Physiol.*, 24, 203 (1940-41).
  - (11) Krejci, L., and Svedberg, T., *J. Am. Chem. Soc.*, 57, 946 (1935).
  - (12) Lamm, O., and Polson, A., *Biochem. J.*, 30, 528 (1936).
  - (13) Quensel, I. B., *Arch. Biochem. Biophys.*, 1, 441 (1938).
  - (14) Watson, C., *Ann. N. Y. Acad. Sci.*, 37, 322 (1938) and Unpublished.
  - (15) Svedberg, T., *Bull.*, 66, 191 (1934).
  - (16) Pedersen, K. O., *Biochem. Z.*, 285, 207 (1936).
  - (17) Philpot, J. St. L., and Eriksson-Quensel, I. B., *Nature*, 132, 351 (1933).
  - (18) Svedberg, T., *Tab. Biol.*, 11, 351 (1936).
  - (19) Eriksson-Quensel, I. B., *Tab. Biol.*, 11, 351 (1936).
  - (20) Sjögren, B., and Svedberg, T., *J. Am. Chem. Soc.*, 53, 2657 (1931).
  - (21) Svedberg, T., and Sjögren, B., *J. Am. Chem. Soc.*, 51, 3594 (1929).
  - (22) Svedberg, T., and Sjögren, B., *J. Am. Chem. Soc.*, 50, 525 (1928).
  - (23) Pedersen, K. O., *Biochem. Z.*, 285, 207 (1936).
  - (24) Sumner, J. B., *J. Biol. Chem.*, 125, 45 (1938).
  - (25) Gralén, N., *Biochem. Z.*, 285, 207 (1936).
  - (26) Van Dyke, H. B., Snow, B. F., Greep, R. O., and Rothen, A., *Am. J. Physiol.*, 133, 473 (1941).
  - (27) Svedberg, T., and Fähræus, R., *J. Am. Chem. Soc.*, 48, 430 (1926).
  - (28) Tiselius, A., and Gross, D., *Kolloid Z.*, 66, 11 (1934).
  - (29) Svedberg, T., and Sjögren, B., *J. Am. Chem. Soc.*, 50, 3318 (1928).
  - (30) Kekwick, R. A., *Biochem. J.*, 32, 552 (1938).
  - (31) Kekwick, R. A., and Pedersen, K. O., *Biochem. J.*, 30, 2201 (1936).
  - (32) Petermann, M. L., and Papenberger, A. M., *J. Physical Chem.*, 45, 1 (1941); see also Tiselius, A., and Dahl, O., *Arch. Biochem. Biophys.*, 14, 317 (1941).
  - (33) Norrman, J. H., *J. Gen. Physiol.*, 25, 45 (1941-42).
  - (34) Butler, A., *J. Gen. Physiol.*, 25, 47 (1941-42).
  - (35) Mutzenbecher, P. v., *Biochem. Z.*, 266, 250 (1933).
  - (36) Kabat, E. A., *J. Exptl. Med.*, 69, 103 (1939).
  - (37) Eriksson-Quensel, I. B., *Biochem. J.*, 32, 585 (1938).
  - (38) Gralén, N., *Biochem. J.*, 33, 1342 (1939).
  - (39) Svedberg, T., and Sumner, J. B., *J. Am. Chem. Soc.*, 51, 2170 (1929).
  - (40) Svedberg, T., and Sjögren, B., *J. Am. Chem. Soc.*, 52, 270 (1930).
  - (41) Polson, A., *Nature*, 137, 749 (1936).
  - (42) Lamm, O., *Nova Acta Reg. Soc. Scient. Upsal.* IV, 10, No. 8 (1937).
  - (43) Sumner, J. B., and Gralén, N., *J. Biol. Chem.*, 125, 33 (1938).
  - (44) Svedberg, T., and Andersson, K. J. I., Unpublished.
  - (45) Eriksson-Quensel, I. B., and Sumner, J. B., *J. Biol. Chem.*, 125, 37 (1938).
  - (46) Sumner, J. B., Gralén, N., and Eriksson-Quensel, I. B., *J. Biol. Chem.*, 19, 95 (1935).
  - (47) Heidelberger, M., and Pedersen, K. O., *Science*, 87, 322 (1935).
  - (48) Kabat, E. A., and Pedersen, K. O., *Science*, 87, 322 (1935).
  - (49) Svedberg, T., and Eriksson, I. B., *J. Am. Chem. Soc.*, 54, 4730 (1932).
  - (50) Carter, R. O., *J. Am. Chem. Soc.*, 63, 1960 (1941).
  - (51) Svedberg, T., and Eriksson, I. B., *J. Am. Chem. Soc.*, 55, 2834 (1933).
  - (52) Svedberg, T., and Chirnoaga, E., *J. Am. Chem. Soc.*, 50, 1399 (1928).
  - (53) McFarlane, A. S., and Kekwick, R. A., *Biochem. J.*, 32, 1607 (1938).
  - (54) Lauffer, M. A., *J. Physical Chem.*, 44, 1137 (1940).
  - (55) Lauffer, M. A., and Stanley, W. M., *J. Biol. Chem.*, 135, 463 (1940).
  - (56) Neurath, H., Cooper, G. R., Sharp, D. G., Taylor, A. R., Beard, D., and Beard, J. W., *J. Biol. Chem.*, 140, 293 (1941).
  - (57) McCalla, A. G., and Gralén, N., *Canadian J. Res.*, 20, 130 (1942).
  - (58) Neurath, H., and Cooper, G. R., *J. Biol. Chem.*, 135, 455 (1940).

1. Molecular Weight Classes. Svedberg (reference 1, p. 406) has concluded that the molecular weights found are not distributed at random, but fall into groups the weight of the molecules in any one of which



in the 35,200 ( $= 2 \times 17,600$ ) group. The weights actually estimated for molecules in this class range from 30,000 for crotoxin to more than 40,000 for egg albumin and concanavalin B. Thus, if there is a factor truly common to all the molecules in this group, it cannot be the molecular weight as such; it is conceivable, however, that all these molecules consist of the same number of amino acid residues. This would imply that the average amino acid residue weight is very different for the different proteins—a question to be decided by analytical studies. The studies of Bergmann and Niemann (Chapter 13) have indeed given tentative indications that this is the case for certain proteins, but the problem is still by no means settled.

Recently two highly purified proteins have been shown to have molecular weights near 13,000—namely, ribonuclease<sup>24</sup> and cytochrome C<sup>25</sup>. The higher value given in Table 1 for cytochrome C was obtained on less pure material. The difference between 13,000 and 17,600 is definitely beyond experimental error, so that these two proteins fail to fit into Svedberg's scheme<sup>26</sup>.

Svedberg<sup>1</sup> has estimated the probable error in the estimated diffusion and sedimentation constants as 2 to 3%, and in the calculated molecular weights as 5 to 10%. Thus the data certainly do not permit any inference that the larger protein molecules (with molecular weights of 400,000 or more) are constructed of units with a weight of 17,600—the probable error is far too great for this. Some very large units, such as those of the hemocyanins, do, however, break up readily, and often reversibly, into smaller units (see below under "Dissociation and Association Reactions").

The molecular weights of the oxygen transporting blood proteins studied show a definite correlation with their biological environment. All those of low molecular weight (with one exception) are enclosed in corpuscles, while those which are carried in solution in the plasma are distinguished by high molecular weights and sedimentation constants. The blood hemoglobins of all the vertebrates, except for the lowest class, the *cyclostomata*, are of the same molecular weight (about 68,000) containing four iron atoms per molecule. The molecular weight of the hemoglobin of the *cyclostomata* is only a quarter of this; and certain invertebrates have corpuscle erythrocruorins of half the molecular weight of hemoglobin<sup>27</sup>. Erythrocruorins of high molecular weight occur in the blood plasma of the crustacean *Daphnia* ( $M = 400,000$ ), the snail *Planorbis* ( $M = 1,600,000$ ), and certain worms such as *Arenicola* and *Lumbricus* ( $M = 3,200,000$ ). The hemocyanins, which are always found free in solution in the plasma, have

molecular weights which range from about 400,000 (for *Pandalus* and *Palinurus*) to 6,600,000 (for the principal component in the blood of *Helix Pomatia*).

Heidelberger, Pedersen and Kabat<sup>28</sup> studied certain antipneumococcus antibodies by sedimentation and diffusion measurements. They reported the antibodies of pig, cow, and horse to be very large, with sedimentation constants ( $s_{20}$ ) near 18 *S* (molecular weight near 990,000), while those of monkey, man, and rabbit were reported to be much smaller, giving values near 7 *S* (molecular weight near 160,000), close to those characteristic of the major component of normal serum globulin. Both types of molecule are markedly asymmetrical, the larger antibodies showing marked double refraction of flow (Chapter 21)<sup>29</sup>. For both types of antibody, Kabat<sup>28c</sup> found a considerable variation in sedimentation constant with concentration.

The observations of Petermann and Pappenheimer<sup>29a</sup> on antipneumococcus antibodies throw additional light on the problem. The antibody euglobulin from Type I serum was concentrated in a component with  $s_{20} = 7$  *S*; while a euglobulin fraction from Type II antiserum showed major components with  $s_{20} = 7, 11, 18$  and 30 *S*. In the latter case the antibody was concentrated mainly in the three heaviest components. Thus it would appear that, even within a single species, a particular type of antibody activity is not necessarily associated with a particular size of molecule.

In ultracentrifugal measurements on globulin fractions from normal horse serum, Oncley<sup>29b</sup> found that  $\alpha$ -pseudoglobulin contains approximately equal proportions of material sedimenting, in a one per cent solution, with  $s_{20}$  values of 17, 6.8, and 3.6 *S*. The first of these components is comparable in size to the large antibody molecules; the second shows a typical value for serum globulin; the third sediments even more slowly than normal serum albumin. Evidence for the presence of more than one component was found also in some of the other globulin fractions studied in these investigations. It would appear, therefore, that very large globulin molecules may appear in normal serum and are not necessarily antibodies. It is also possible, although not yet proven, that a reversible equilibrium exists between the large and small globulin components (compare the discussion on p. 437-443 below).

**2. Comparison with Osmotic Data.** Molecular weights calculated from osmotic pressure data (Chapter 17) show on the whole a very satisfactory

<sup>28</sup> (a) Heidelberger, M., and Pedersen, K. O., *J. Exptl. Med.*, 65, 393 (1937); (b) Heidelberger, M., and Kabat, E. A., *J. Exptl. Med.*, 67, 181 (1938); (c) Kabat, E. A., *J. Exptl. Med.*, 69, 103 (1939).

agreement with ultracentrifugal measurements. Such discrepancies as are found may probably be attributed largely to polydispersity of the solutes used for the osmotic pressure studies. Also the accuracy of molecular weights calculated from osmotic pressure is low for solutes of high molecular weight, and ultracentrifugal values are far more accurate for such materials. It is of course obvious that molecular weights obtained in water are generally not comparable with those obtained in such solvents as concentrated urea solutions, since many proteins dissociate into smaller components in the latter solvent. However, serum albumin and the hemoglobins of certain species appear to have the same molecular weights in both solvents.

**3. Frictional Ratios, Molecular Shape, and Hydration.** An unhydrated spherical molecule should give a frictional ratio of unity. It may be seen from the values in Table 1, however, that all the proteins hitherto studied by sedimentation and diffusion measurements give frictional ratios appreciably greater than 1. It may be seen from Fig. 1 that a molecule which carries 0.3 gram of water of hydration per gram of protein should give a frictional ratio of approximately 1.13. This is an entirely plausible value for the possible hydration, considered in the light of the work of Adair and Adair on the density of protein crystals (Chapter 16). A number of the proteins studied have frictional ratios lower than 1.13 and these molecules may conceivably be spherical. Many of the proteins studied, however, give much higher frictional ratios—see, for instance, zein and the large antibody globulins of horse sera—which may be higher than 1.5 and even as great as 2. In such molecules the observed frictional ratios cannot be explained by hydration alone, unless impossibly large hydration values are assumed. We are forced to conclude, therefore, that these molecules deviate very markedly from the spherical shape. This conclusion is amply confirmed by studies of viscosity and double refraction of flow (Chapter 21), and of the dispersion of the dielectric constant (Chapter 22).

Sedimentation and diffusion measurements, while they thus establish the existence of non-spherical proteins, cannot prove whether these proteins are rod-shaped or disc-shaped molecules, or whether perhaps their structure corresponds to some more complicated geometrical pattern. Polson<sup>30</sup> and Neurath<sup>31</sup> have estimated the shapes of protein molecules, assuming that for purposes of calculation they may be treated as ellipsoids of revolution, and have calculated the axial ratio,  $a/b$ , employing Perrin's equation (Chapter 18, equations 27 and 28). Assuming an elongated (prolate) ellipsoid, and zero hydration, zein is thus calculated to have a major axis of 322 Å. and a minor axis of 16 Å. Neurath's calculations

indicated a minor axis of at least 18 Å for all the other proteins for which data were available. These represent minimum values for the minor axis, since any correction for hydration reduces the calculated axial ratio,  $a/b$ , and thus increases the calculated length of the minor axis. Tobacco mosaic virus nucleoprotein, which is a very large rod-shaped molecule, has been shown from x-ray diffraction studies (Chapter 14) and from electron microscopy (Chapter 21) to have a diameter of 150 Å.

Thus even highly asymmetrical protein molecules have a very considerable thickness, and are profoundly different in structure from the thread-like synthetic polymers such as polystyrene, polyacrylic acid, polyoxymethylene, and the poly- $\omega$ -hydroxydodecanoic acids<sup>32, 33</sup>. Such polymers are only 3 to 6 Å in diameter, in the direction perpendicular to the chain length—that is, in this dimension they are of size similar to that of the solvent molecules, although extremely elongated along the length of the chain. Protein molecules, on the other hand, are large in all dimensions compared to the molecules of the solvent. Furthermore, their relaxation times, as estimated from dielectric dispersion measurements (Chapter 22), indicate that a protein molecule orients as a whole in an electric field; that is, the protein is an essentially rigid structure, in contrast to the flexible molecules of the synthetic high polymers. Extremely elongated protein molecules such as myosin, which give very viscous solutions and show intense double refraction of flow (Chapter 21) may be thinner in cross-section, and thus more flexible, than most proteins. It is not yet proved, however, that even the myosin molecule can bend or coil easily; and certainly most native proteins have relatively rigid structures.

**4. Effects of Denaturing Agents.** We have already seen, from the work of Neurath and Saum (Chapter 18, Table 6) that serum albumin, in concentrated urea solution, appears to uncoil into a relatively elongated form, as judged by the change in its diffusion constant. Similar effects appear to exist in other proteins. Mirsky<sup>34</sup> has obtained indirect evidence, from the titration of the sulfhydryl groups liberated on the denaturation of egg albumin, that the denaturation process in this protein involves an unfolding of the molecule. Sedimentation studies by Rothen<sup>26</sup> on Mirsky's preparations confirmed these conclusions. The value of  $f/f_0$  in native egg albumin is 1.16 (Table 1); in concentrated urea solution it rises to a value near 4, corresponding to an axial ratio of about 90 to 1, indicating a complete unfolding of the protein. "If the solution is then dialyzed, the molecules refold as judged by titration and sedimentation, but the refolding happens in an apparently random way, the solution is polydisperse, aggre-

gation has occurred" (reference 26, p. 240). In collaboration with Landsteiner, Rothen found that if egg albumin is heat denatured in the complete absence of salt, unfolding occurs in a definite way, the solution is monodisperse, and  $s_{20} = 7.0$  Svedberg units, a value twice as great as that of native albumin. If the heat denaturation is carried out in the presence of 0.2% NaCl, the solution is also monodisperse but  $s_{20} = 17$  Svedberg units, as determined in the presence of 1% NaCl after cooling.

Studies of surface films of proteins<sup>35, 36</sup> reveal a thickness of the protein monolayers of only about 7 Å. Since this is much smaller than any of the dimensions we have seen to be characteristic of protein molecules in solution, it appears that a radical change must take place in the structure of a native protein before it can spread on a surface. This change also must be in the nature of an unfolding, though whether the character of the unfolding is the same in surface denaturation, heat denaturation and urea denaturation must for the present remain undecided.

The work of Burk, showing the dissociation of edestin, amandin, excelsin and horse hemoglobin in concentrated urea, as judged by osmotic pressure measurements, has already been discussed in Chapter 17. Burk's work on horse hemoglobin has been confirmed and extended by the ultracentrifugal and diffusion measurements of Steinhardt<sup>36a</sup>, who has found the molecular weight in urea solution above 4 molar to be about 38,000. This is little more than half the molecular weight in water; it may be taken as half practically within the limits of experimental error. This dissociation is at least partially reversible by removal of the urea by dialysis. The oxygen capacity and absorption spectrum of hemoglobin were found by Steinhardt to be the same in concentrated urea as in water, so the dissociation process taking place in this solvent apparently involves a much less radical change in the molecular pattern than does the denaturation produced by heat, acids and alkalies. There was also little change in the frictional ratio, indicating that there is much less change of shape in the hemoglobin, due to the action of urea, than there is in egg or serum albumin. Acetamide was found to produce effects on hemoglobin similar to those of urea but only at somewhat higher concentrations. The molecular weight of pepsin was the same in 4 molar urea as in water, and the protein retained its enzymic activity in this solvent.

Anson and Mirsky<sup>37</sup> concluded that the denaturation of hemoglobin is completely reversible under suitable conditions. The same conclusion was reached for crystalline trypsin by Northrop<sup>38</sup> and by Anson and Mirsky<sup>39</sup>

<sup>35</sup> Langmuir, I., *Cold Spring Harbor Symp. Quant. Biol.*, 6, 171 (1938).

<sup>36</sup> Gorter, E., in Schmidt, C. L. A., "*Chemistry of Amino Acids and Proteins*." C. C. Thomas Springfield

Recently Gralén<sup>40</sup>, following the procedure of Anson and Mirsky, split hemoglobin into heme and globin by the use of acetone and dilute acid, and then combined the heme and globin to obtain a new product. Anson and Mirsky had concluded that this product was identical with the original native hemoglobin, and Gralén has shown that it is monodisperse and has the same molecular weight (69,000) as native hemoglobin, the same pH stability diagram, and the same molecular shape. It differs somewhat, however, from the original hemoglobin in electrophoretic mobility, so that complete identity with the original hemoglobin cannot be assumed.

The recent work of Neurath, Cooper and Erickson<sup>40a</sup> on the albumin and pseudoglobulin of horse serum has shown that the alterations produced in these proteins, by exposure to concentrated solutions of urea or of guanidine hydrochloride, are only in part reversible. Removal of the denaturing agent by dialysis causes a separation of the protein into two fractions. The more insoluble fraction is radically different from the original material; the more soluble fraction is generally similar to the undenatured protein, but viscosity and diffusion measurements indicate that it is by no means identical. The denaturing effect of guanidine hydrochloride is greater than that of urea at any given concentration. This finding is in accord with the measurements of Greenstein<sup>40b</sup>, which showed guanidine hydrochloride to be much more effective than urea in increasing the titratable sulfhydryl groups of many proteins.

The uncoiling of native proteins into a more elongated form, due to the action of denaturing agents, has been inferred from x-ray studies by Astbury, Dickinson and Bailey (Chapter 14). The results of these sedimentation and diffusion measurements on denatured egg and serum albumin are entirely in harmony with their conclusions. It must not be inferred, however, that the effects of these reagents are the same for all proteins. The action of urea and guanidine hydrochloride on the properties of myosin (Chapter 21) is radically different from its effect on the albumins. The action of denaturing agents on the size and shape of protein molecules may vary with the geometry of the particular molecule concerned, and with the arrangement of the labile linkages, within the native protein molecule, on which the denaturing agent presumably acts<sup>41</sup>.

**5. Dissociation and Association of Protein Molecules.** The molecular weight of a protein is a definite quantity under well-defined conditions of

<sup>40</sup> Gralén, N. *Hingdon J.* 33, 1907 (1939).  
<sup>40a</sup> Neurath, H., and Cooper, R. A., and Erickson, F. O. *J. Biol. Chem.*, 142, 249, 265 (1942).  
<sup>40b</sup> Greenstein, J. P. *J. Biol. Chem.*, 128, 233 (1939); 130, 519 (1939); Greenstein, J. P., and Edsall, L. S. *J. Biol. Chem.*, 144, 1 (1944).  
<sup>41</sup> Neurath, H. *J. Biol. Chem.*, 22, 439 (1936), have advocated the view that these labile linkages are hydrogen bonds, which can certainly be formed between the  $-C=O$  and  $-NH$  groups in

pH and temperature, and in solvents in which the protein is stable. Relatively small alterations in the environment, however, may cause such a molecule to break up into smaller units, or to associate into larger ones. These alterations were often not clearly recognized by chemists before the advent of the ultracentrifugal technique; they do not involve such gross changes in the properties of the protein as do the types of denaturation discussed above. Nevertheless, the two groups of phenomena are closely connected.

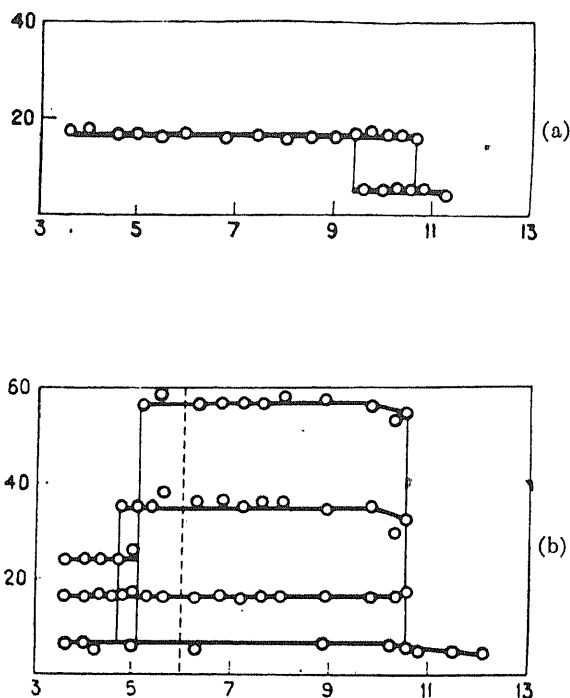


Figure 2. pH-stability diagrams for hemocyanins, (a), *palinurus vulgaris*, and (b), *limulus polyphemus*. Abscissae in both figures, pH; ordinates,  $s_{20}$ . The dotted line in (b) indicates the positions of the isoelectric points. From Eriksson-Quensel, I.-B., and Svedberg, T., *Biol. Bull.*, 71, 498 (1936).

Alterations of the sedimentation and diffusion constants, arising from pH changes, are particularly striking in many of the hemocyanins, which have been studied in great detail by Eriksson-Quensel and Svedberg<sup>42</sup>. A simple instance of this is the hemocyanin of the crustacean, *Palinurus Vulgaris*, of the order *Malacostraca* (Fig. 2). In the pH region 3.6 to 9.4 this hemocyanin consists of a single molecular species. of sedimentation

ponent of sedimentation constant 4.10. The amount of this component increases with pH, and at pH 10.8 the larger component has disappeared.

The hemocyanin of *Limulus Polyphemus* (the horseshoe crab), has four well defined components in the pH range 5.2–10.5 (Fig. 2). Their sedimentation constants are 56.6 (*D*), 34.6 (*F*), 16.1 (*H*) and 5.87 (*K*); and the relative proportions of these components are unchanged within this pH region. "Above pH 10.5 *D*, *F* and *H* disappear; they split up into *K*. Below pH 5.2 *D* and *F* dissociate in two ways. The concentration of *K* apparently increases; *i.e.*, some part of them must form more of *K*. A new component, *G*, with the sedimentation constant 24.0 appears in rather high concentration; and *G* must be a dissociation product of *D* and *F*, as the concentration of *H* does not change and that of *K* increases" (reference 42, p. 511).

These dissociation reactions are completely reversible. If the solution is brought to pH 10.8 or to 4.0, and then brought back to the intermediate pH range, the four original components are again found, in their original proportions (reference 42, p. 538).

Another type of behavior is shown by the hemocyanin of the snail, *Helix Pomatia* (Fig. 3). This "appears in four different well-defined molecular species, *B* with the sedimentation constant 98.9, *C* with 62.0, *H* with 16.0 and *I* with 12.1. The main component is *B*, present in solutions with a pH from 3.6 to 8.2. It is alone in the pH region 4.6–7.4 and accompanied by *C* in the regions 3.6–4.6 and 7.4–8.2. *H* is present in very small quantity together with *B* and *C* in the pH range 7.9–8.2, alone between pH 8.2 and 9.3, and together with *I* in the pH range 9.3–10.8. In more alkaline solutions the hemocyanin is inhomogeneous and the sedimentation constant decreases. Below pH 3.6 the material is non-uniform and the sedimentation constants are irregular but apparently decreasing" (reference 42, pp. 519–521).

The dissociation is reversible inside a certain pH range (up to about pH 9), but if the solution is made too alkaline only a part of the hemocyanin is reassociated into the normal components on returning to a pH near 6.

On the basis of these and a very large number of other data, Eriksson-Quensel and Svedberg conclude (reference 42, p. 541–43): "The most striking result of this investigation is perhaps the perfect homogeneity of the various hemocyanins with regard to molecular weight. Not only the main component or components which exist in the region comprising the isoelectric point, but also the products of reversible dissociation, are definite molecular individuals. At very low and very high pH values where irre-



the isoelectric point. . . . The molecular weights of the hemocyanin found in the blood of a certain species are always simple multiples of the lowest well defined component. Thus for the *Malacostraca* the relationship is 1:2 and for the *Gastropoda* 2:8:16:24." All species belonging to the same order show similar pH stability diagrams.

These reversible dissociation reactions can be brought about by very small pH changes, which would certainly have little influence on the structure of most molecules. Other factors than pH can also influence these reactions profoundly. Thus the presence of divalent cations ( $\text{Ca}^{++}$ ,  $\text{Mg}^{++}$ )

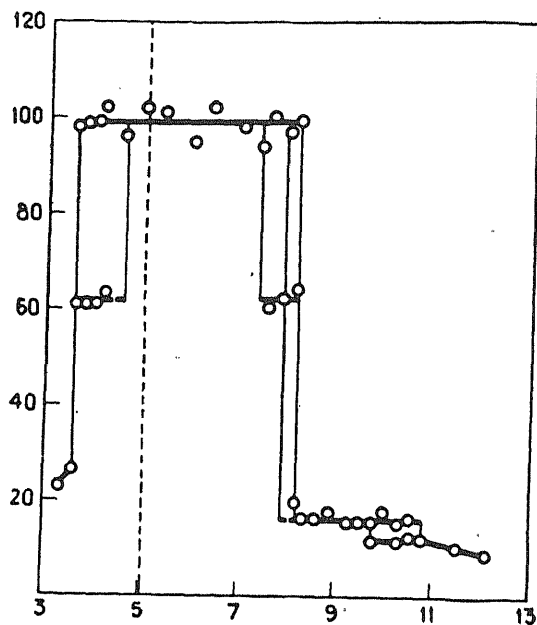


Figure 3. pH-stability diagram for *helix pomatia*. Abscissa, pH; ordinate,  $s_{20}$ . The dotted line indicates the position of the isoelectric point. From Eriksson-Quensel, I.-B., and Svedberg, T., *Biol. Bull.*, 71, 498 (1936).

profoundly influences the pH stability diagram of the hemocyanin of *Helix Pomatia*. The first indication of such an effect was obtained from a study of the Tyndall effect in these solutions<sup>43</sup>. Ultracentrifugal analysis has shown that in the presence of 0.01 molar calcium chloride, dissociation on the alkaline side of the isoelectric point does not become apparent until pH 9.5 is reached; while in the absence of calcium this change begins to appear at pH 7.4<sup>44</sup>. The splitting of the hemocyanin molecules may also

be produced by ultrasonic waves<sup>45,46</sup> and by ultraviolet light<sup>47</sup>. The effects of the two latter agents are probably irreversible

If hemocyanins from two different but closely related species are mixed, brought to a pH where dissociation of both occurs, and then brought back to a pH near the isoelectric point, mixed molecules are formed, containing components from both of the original dissociated types. This has been shown for the hemocyanins of *Helix Pomatia* and *Helix Nemoralis* by Tiselius and Horsfall<sup>48</sup>. The two original hemocyanins, and the mixed reassociated products, all have nearly the same molecular weights, but they may be distinguished by their different electrophoretic mobilities. These authors conclude that "the various mixed hemocyanins which have been produced are probably as stable as the hemocyanins from which they were formed" (reference 48, p. 100). The hemocyanins of *Helix Pomatia* and *Littorina Littorea*, which are more distantly related biologically, also form mixed molecules, but less readily and with the formation of a number of components of differing molecular weight.

Horse CO-hemoglobin, in moderately concentrated salt solutions (0.5–1.0 molar) is dissociated into molecules having half the normal weight. Further increase in salt concentration produced no further change<sup>49</sup>.

Dissociation of protein molecules can be produced by the presence of other proteins, or of certain amino acids and other amino compounds. Interaction between proteins is particularly marked in the serum proteins, a phenomenon first revealed by the work of McFarlane<sup>50</sup>. In concentrated serum, part of the globulin splits up into smaller molecules, an effect which is probably produced by serum albumin or one of its components<sup>51</sup>. The effect differs for different species<sup>51a</sup>. When diluted with seven volumes of saline solution, the sera of horse, cow and man all show the normal sedimentation constants of serum albumin and serum globulin. The globulin content of the three species decreases in the order named. In the undiluted sera the amount of material sedimenting with the characteristic globulin velocity is much less. In one case (human serum) a new component appears, possibly corresponding to half molecules of globulin. In horse and cow serum, the sedimentation constant of this component probably coincides with that of albumin. Addition of salts, to human serum splits the dissociated globulin still further, to two products with molecular

<sup>45</sup> Brändén, S., *Natur.*, **110**, 865 (1937).

<sup>46</sup> Brändén, S., *Nord. Act. Soc. Scient. Upsaliensis*, **4**, 12, No. 9 (1940).

<sup>47</sup> Svedberg, T., and Brändén, S., *Nature*, **142**, 830 (1938).

<sup>48</sup> Tiselius, A., and Horsfall, J. L., Jr., *J. Exptl. Med.*, **69**, 83 (1939).

<sup>49</sup> Unpublished studies of Pedersen and Andersson, cited by Pedersen (reference 1, p. 407).

<sup>50</sup> McFarlane, A. S., *Biochem. J.*, **29**, 660 (1935).

<sup>51</sup> Pedersen, P. O., *Acta Med. Scand.*, cited by Svedberg, T., *Kolloid Z.*, **68**, 192 (1929).

weights of the order of 40,000 and 20,000. Similar studies have recently been made on the sera of birds, reptiles, amphibia and fishes.<sup>52</sup>

Such dissociation reactions, due to the interaction between proteins, or to the dilution of a single protein component, have been extensively studied by Pedersen<sup>53</sup>. Recently<sup>54</sup> he has offered the suggestion that the primary units of protein structure (presumably composed of polypeptide chains somehow fixed in a definite pattern) are held together by a non-protein "cement", which may consist of carbohydrates, phosphoric ester groups, nucleic acids, or other residues. One protein may partially effect

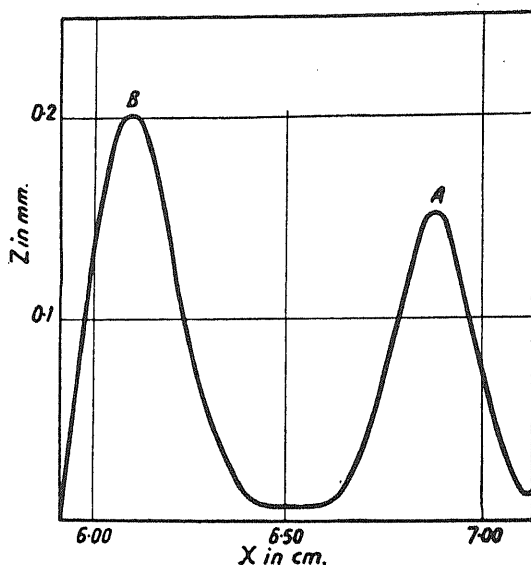


Figure 4. Sedimentation diagram of serum albumin in solution of 2.6% clupein; the main maximum with rapid sedimentation, *A*, corresponds to undissociated protein, the maximum of slower sedimentation, *B*, to the dissociation product with  $s \sim 1$  and  $M \sim \frac{1}{2}$  of the serum albumin; the sedimentation of the clupein itself has been deducted from the curves. Time: 100 minutes after full speed was reached. From K. O. Pedersen, in Svedberg, T., *Kolloid-Z.*, **85**, 119 (1938).

the transfer of this cement from another to itself, thus influencing the stability of the other. At present there seems to be insufficient evidence to prove or disprove this conception, but it affords a highly suggestive basis for future work.

Salts of some of the simple protamines, such as clupein, cause the partial dissociation of serum albumin (reference 1, p. 408) (Fig. 4). There is a high degree of specificity in the action of dissociating agents on proteins.

Thus serum albumin is split (at neutral pH) by arginine + ammonium chloride, but not by lysine + ammonium chloride; the exact reverse is true for *Helix* hemocyanin. Guanidine hydrochloride has a very powerful effect on *Helix* hemocyanin, but very little on serum albumin. Clupein splits both, while arginine without ammonium chloride has no action on either<sup>8c</sup>. It remains to be seen how far the action of clupein on proteins is reversible.

These dissociation reactions are among the most important and characteristic reactions of protein molecules. Similar phenomena are revealed in the study of other classes of proteins by quite different methods. Thus the changes produced by the above mentioned reagents, and by many others, on viscosity and double refraction of flow in myosin (Chapter 21) may rest on a very similar underlying basis to those revealed by the ultracentrifugal studies on blood proteins. It appears clear that the components of the protein molecule are linked together largely by very strong peptide bonds, but that there are also a considerable number of much weaker bonds found in many native proteins; the latter are readily split by extremely mild reagents. Such bonds may be peptide linkages which have been rendered unstable owing to the particular configuration of the adjoining groups; they may involve carbohydrate or other prosthetic groups as suggested by Pedersen; and they may involve salt linkages between charged groups, or hydrogen bonds. The study of these reactions represents one of the most important fields in all of protein chemistry.

## Chapter 20

### Proteins as Acids and Bases

BY JOHN T. EDSALL

The amino acids of which proteins are composed contain several different types of acid and conjugate basic groups. In accordance with the terminology of Brönsted, (Chapter 4) the acid groups may be divided into two different classes. The first is represented by the uncharged acid groups such as the carboxyl, sulfhydryl, and phenolic hydroxyl groups, which carry no net charge, in strongly acid solution, the conjugate basic groups being negatively charged in alkaline solution. The second class is represented by the cationic groups, such as the guanidinium group in arginine, the imidazolium group in histidine, and the  $\epsilon$ -ammonium group in lysine. These groups are positively charged in acid solution and the conjugate basic groups carry no charge in alkaline solution.

The peptide linkage, the amide groups of asparagine and glutamine and the aliphatic hydroxyl groups of serine, threonine, hydroxyproline and hydroxyglutamic acid, do not function as either acids or bases within the pH range 1-13, within which protein titrations have generally been conducted.

Since almost all the  $\alpha$ -ammonium and  $\alpha$ -carboxyl groups are bound in peptide linkage, these in general play no significant part in acid and base binding by proteins. Occasional groups of this sort may be free at the end of long peptide chains, and contribute a very small fraction of the total acid and base binding. It should be noted, however, that cystine may play a special part in this respect since only two of its four acid and basic groups are necessarily bound in peptide linkage, although all four may be. If only two are bound, the other two which are also  $\alpha$ -amino and carboxyl groups may be free to function in combining with acids and bases. In general, however, the titratable carboxyl groups in proteins are the free terminal groups of aspartic, glutamic, and hydroxyglutamic acids, not bound as amides; and the titratable ammonium groups are the  $\epsilon$ -ammonium groups of lysine. The characteristic pK values of all these groups have already been discussed in detail in Chapter 15.

range of pH within which these groups dissociate is significantly altered by adjoining charged groups or polar groups in a more complex molecule such as a protein. However, if both the pK value of a certain group in a protein and its heat of ionization be known, little doubt generally remains as to the nature of the group.

In the present chapter we shall consider only the uptake or giving off of protons. A clear distinction exists between the binding of protons by basic groups, which involves the formation of a true covalent bond, and the possible binding of other ions, which involves only an electrovalent bond. While ions other than the proton may sometimes be held to acid or basic groups by relatively strong forces, particularly if the ions involved are polyvalent, still the bond so formed is of a different character from the bond formed by the attachment of a proton to a basic group. In the

TABLE 1. CHARACTERISTIC ACIDITY CONSTANTS (EXPRESSED AS pK VALUES) AND HEATS OF IONIZATION ( $\Delta H$ ) OF ACIDIC AND BASIC GROUPS FOUND IN PROTEINS

Group	pK (25°)	$\Delta H$ (cal/mol)
Carboxyl ( $\alpha$ ).....	3.0- 3.2	( $\pm 1500$ )
Carboxyl (aspartyl).....	3.0- 4.7	( $\pm 1500$ )
Carboxyl (glutamyl).....	ca. 4.4	( $\pm 1500$ )
Phenolic-hydroxyl (diiodotyrosine).....	6.5	800
Phenolic-hydroxyl (tyrosine).....	9.8-10.4	6000
Sulfhydryl.....	9.1-10.8	
Imidazolium (histidine).....	5.6- 7.0	6900-7500
Ammonium ( $\alpha$ ).....	7.6- 8.4	10,000-13,000
Ammonium ( $\alpha$ , cystine).....	6.5- 8.5	
Ammonium ( $\epsilon$ , lysine).....	9.4-10.6	10,000-12,000
Guanidinium (arginine).....	11.6-12.6	12,000-13,000

The pK values given are those characteristic for the groups named, in peptides of known structure (see Chapter 4, Table 5, for complete details). The value for the —OH group in diiodotyrosine is for the free amino acid, since no peptides containing it have been studied.

The  $\Delta H$  values are derived from Tables 2 and 6 of Chapter 4, and are for both amino acids and peptides.

discussion in this chapter we shall consider the charge on the protein molecule as determined by the hydrogen ion activity in the system and the ionization constants of the acid and basic groups in the protein molecule.

In a strongly acid solution all acid groups in the protein molecule which are of the uncharged type carry zero charge. All cationic groups carry positive charges. Such a condition is actually attained, to a very close approximation, between pH 1 and 2, since the most acid carboxyl groups in the protein molecule are characterized by pK values somewhat greater than 3. In a strongly alkaline solution, on the other hand, all the uncharged acid groups have lost protons and carry a negative charge. All the cationic acid groups have lost their protons and carry a zero charge.

tions which have been carried out on proteins; since even at this pH an appreciable number of the guanidino groups of the arginine residues may still retain a positive charge. Thus the acid bound by proteins may be expected to attain a maximum value, independent of pH, between pH 1 and 2; but no similar sharply defined maximum in base binding is to be expected, even between pH 12 and 13. On the whole, these expectations appear to be borne out by the experimental facts to be reported in this chapter.

### Net Charge, Total Charge and Maximum Acid and Base Binding

If the titration of a protein is begun in a strongly acid solution in which the protein molecule carries its maximum positive charge, the addition of a strong base (generally hydroxyl ion) brings about the removal of protons from the acidic groups on the protein. If a proton is removed from an uncharged acid group, a negatively charged group is produced; thus the *total* charge on the protein is increased by one unit while the *net* charge is decreased by the same amount. If a proton is removed from a cationic acid group, both the net charge and the total charge decrease by one unit. Thus, if the protein contains  $n$  cationic acid groups, the net (and also the total) charge in strongly acid solution is  $+n$ . When  $n$  protons are removed from the protein ion by the addition of base, the net charge on the protein molecule falls from  $+n$  to zero, although the total charge on the protein in this condition may be, and generally is, much greater than on the cationic protein. The pH of the protein solution after these  $n$  protons have been removed by addition of base is known as the isoionic point. If the protein binds no other ions than protons, the isoionic point may correspond to the isoelectric point of the protein determined by electrophoresis. If the protein combines with other ions also, the isoelectric and isoionic points are different. In the following discussion, however, we shall generally make no distinction between isoelectric and isoionic points<sup>1</sup>.

The increase in net charge on the protein molecule between the isoionic point and the region of maximum acid binding capacity should clearly be equal to the number,  $n$ , of cationic acid groups in the protein. This is apparent from the argument given above and is true independently of the fact that most of the groups which acquire protons on the addition of acid are not the amino, guanidino, or imidazole groups, but are chiefly the negatively charged carboxylate ions.

By exactly similar reasoning it is clear that the maximum base binding capacity of the protein, if it can be determined, should be equal to the

number of uncharged acid groups in the protein molecule. This relation renders significant the study of the composition of proteins as determined by analytical methods in relation to their acid and base binding capacity<sup>2</sup>.

### Determination of Nature of Ionizing Groups in Proteins

Several different methods are available for determining the nature of the groups in the protein molecule ionizing in various pH zones. Here we may summarize the methods; the remainder of the chapter will illustrate their application.

1. The pK values given in Table 1 afford an approximate indication of the region in which the different types of acid groups in the protein molecule may be expected to yield up protons to base. These regions may be shifted owing to interaction of the acid group involved with other neighboring groups within the molecule, but, in general, such shifts are relatively small and even in the most extreme cases known are scarcely ever greater than 2 pH units. However, some of the groups which from analytical data might be expected to be available in the protein molecule as proton donors may be tied off, for example, in ester linkage, or in some way inactivated so that they do not contribute titratable groups. Several actual cases are known where there are indications that this may be true.

2. The characteristic heats of ionization of the different groups given in Table 1 afford a very important clue to their nature. By carrying out the titration curve of a protein at two different temperatures, the heats of ionization of the groups involved in different portions of the curve may be determined, and the results so obtained form a valuable supplement to the information deduced by the first method.

3. The addition of small quantities of formaldehyde has little effect upon the titration of carboxylic or phenolic hydroxyl groups, but has a pronounced effect on ammonium and imidazole groups. The shift in titration curve of a protein produced by the addition of formaldehyde may thus indicate quite clearly the region in which the carboxyl groups have become fully ionized and in which the imidazole and ammonium groups begin to yield protons on further addition of base (see also Chapter 5).

4. The ionization of carboxyl groups is affected, by the addition of alcohol, in the opposite sense from the ionization of imidazole, ammonium, and guanidino groups. Thus the titration curve of a protein in alcohol-water mixtures, as compared with its titration curve in pure water, may afford a clue to the nature of the groups ionizing in different pH zones. The effect of alcohol, however, must be interpreted with caution, partly because the alcohol added may denature the protein and partly because



5. Certain specific groups in the protein molecule may be inactivated by chemical treatment. Thus, the  $\epsilon$ -ammonium group of lysine may be converted into a hydroxyl group by treatment with nitrous acid. This hydroxyl group, being aliphatic, will not combine with base over the pH range generally employed in protein titration curves; and, if it may be assumed that the nitrous acid has had no effect on any other group in the protein molecule, the resulting titration curve should be the same as that of the original protein, except for the disappearance of the  $\epsilon$ -ammonium groups in lysine. This technique has already been used on more than one occasion in the study of proteins and appears to yield significant results.

Theoretically, it should be possible to tie off the carboxyl groups of the protein molecule by esterification. To accomplish this in practice, however, would involve drastic treatment which might well lead to denaturation of the protein, and to splitting off of some of the groups of the molecule. However, the acetylation of the ammonium groups and the phenolic hydroxyl groups of the protein molecule by ketene is a simple and relatively mild process, and the study of the titration curve of proteins so treated should lead to interesting conclusions.

6. Chemical substitution in the protein molecule may be so employed as to enhance or diminish the acidity of certain ionizing groups. The shift in the titration curve resulting from such treatment then affords an indication of the pK value of the groups of this type in the original protein. This technique is exemplified by the action of iodine on proteins. Iodine reacts readily with the benzene ring of tyrosine, converting it into diiodotyrosine, in which the pK value of the hydroxyl group is about 6.5 instead of being near 10 as in tyrosine itself. The use of this method also will be exemplified in the discussion below. Here, however, as with the other methods of chemical treatment used to alter the titration curve of proteins, there is the difficulty that the added reagent cannot be considered as strictly specific for a single type of group. Iodine appears to attack not only the benzene ring of tyrosine, but the imidazole ring of histidine and probably other groups in the protein molecule as well<sup>3</sup>. It is also a powerful oxidizing agent, converting  $-\text{SH}$  groups to  $\text{S}-\text{S}$ , and even to higher oxidation products<sup>4</sup>, and probably effecting other types of oxidation in the protein molecule. Therefore the application of this technique has led to results which are in some cases confused and exceedingly difficult to interpret. Nevertheless, it remains a valuable method and will undoubtedly be extended in future by the use of other reagents which are highly specific for certain groups in the protein molecule. It should be of great value, for instance, in the interpretation of protein titration curves

if a reagent were available which would react readily and specifically with the guanidino group of arginine, and thereby reduce its  $pK$  value to a point where all the guanidino groups could readily be titrated. The requirements for such reagents, however, are very exacting. They must not only be specific for the particular group involved, but they must be capable of being introduced into that group at low temperatures and in nearly neutral solution, so that the process may occur with a minimum of damage to the protein.

Apparently no reaction of this type is at present known for the guanidinium group of arginine. It is interesting to note, however, that the  $\epsilon$ -ammonium groups of lysyl residues may readily be transformed into guanidinium groups by the action of O-methylisourea<sup>5</sup>.

#### Titration Constants of Polyvalent Acids and Bases

The analysis of protein titration curves shows that they may commonly be described in terms of a limited number of  $pK$  values. In a protein which contains altogether  $n$  acidic groups, these may be broken up into classes each of which contains a number of groups. Thus if one of these classes contains  $q$  groups, the base bound by members of this class may be described as a function of  $pH$  by the equation

$$pH = pK_q + \log \frac{\alpha_q}{1 - \alpha_q} = pK_q + \log \frac{K_q}{(H^+)} \quad (1)$$

where  $K_q$  is an acidity constant characteristic of the members of the group, and  $\alpha_q$  is the fraction of the total number of groups in the class which have given up protons to base, at any  $pH$ . In other words,  $q\alpha_q$  represents the base bound by the  $q$  groups of this class at any  $pH$ .  $(H^+)$ , the hydrogen ion activity is here taken as equal to the negative antilogarithm of the measured  $pH$ . If  $q = 1$ , equation 1, when  $pH$  is plotted against base bound, represents a curve of the same form, in which  $q$  times as much base is bound at any  $pH$  as for a monovalent acid of the same  $pK$  value. That the titration curves of polyvalent acids may be of this form has been shown on theoretical grounds. In 1926, Simms<sup>6</sup> pointed out that the titration curve of one mole of any polyvalent acid containing  $n$  groups could always be described as if it were the titration curve of a mixture of  $n$  moles of monovalent acids with suitably chosen constants (titration constants).<sup>\*</sup> If the successive dissociation constants of the polyvalent acid are denoted by the symbols  $K_1, K_2 \cdots K_n$  and the titration constants by

<sup>5</sup> Greenstein, J. P., *J. Org. Chem.*, **2**, 480 (1937-38).

$G_1, G_2 \cdots G_n$ , then Simms has shown that these constants are related by the equations

$$\begin{aligned} K_1 &= G_1 + G_2 + \cdots + G_n = \sum_i^i G_i \\ K_1 K_2 &= G_1 G_2 + G_1 G_3 + \cdots + G_2 G_3 \cdots + G_{n-1} G_n = \sum_{i,k}^{ik} G_i G_k \\ K_1 K_2 K_3 &= G_1 G_2 G_3 + G_1 G_2 G_4 + \cdots = \sum_{i,k,l}^{i,k,l} G_i G_k G_l \\ &\dots\dots\dots \\ K_1 K_2 K_3 \cdots K_n &= G_1 G_2 G_3 \cdots G_n \end{aligned} \quad (2)$$

If the  $K$ 's are widely separated ( $K_1 > 1000 K_2, K_2 > 1000 K_3$ , etc.) then  $K_1 = G_1, K_2 = G_2$ , etc., and there is no distinction between titration constants and dissociation constants. A very straightforward and elegant proof of these relations was later formulated by von Muralt<sup>7</sup>. The case considered in equation 1 arises when all the  $G$ 's are equal to each other,  $G_1 = G_2 = \cdots = G_n = G$ . Equations 2 then reduce to the form

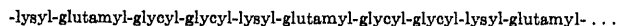
$$K_1 = nG; K_2 = \frac{n-1}{2} G; K_3 = \frac{n-2}{3} G; \cdots K_n = \frac{G}{n} \quad (3)$$

In this case, which has been considered in detail by Weber<sup>8</sup> as well as by von Muralt<sup>7</sup>, the titration curve of the  $n$ -valent acid has the same form as that of  $n$  moles of a monovalent acid having the dissociation constant  $G$ . This condition occurs when a large number of virtually identical acidic groups are part of the same molecule, each of which has the same intrinsic tendency to lose a proton, and each of which is sufficiently far from all the others that the state of ionization of any one group is not influenced by the state of the others.\* Under these conditions the titration constant of the whole set of groups is identical with the intrinsic dissociation constant of any one of the groups in the set, considered individually. In this case, therefore, the titration constant,  $G$ , gives a clearer and simpler indication of the tendency of the groups in the set to dissociate than do the  $n$  dissociation constants  $K_1 \cdots K_n$ . The titration constants also can be calculated from the titration data far more readily than the dissociation constants. In the rest of this chapter, the dissociation constant of an individual acidic group in a polyvalent acid or ampholyte will be denoted

<sup>7</sup> von Muralt, A. L., *J. Am. Chem. Soc.*, 52, 3518 (1930).

<sup>8</sup> Weber, H. H., *Biochem. Z.*, 189, 381 (1927).

\* It is not necessary that other acidic groups of widely different  $pK$  values be present in the groups under consideration. Thus we might consider the case of a polymer containing such a repeating unit as:



Here the lysyl residue and the glutamyl residues would each form a class of the type discussed, each having a characteristic  $pG$  value. The lysyl residues would all carry positive charges in the pH range in which the  $-\text{COOH}$  groups of the glutamyl residues are ionized.

by a small  $k$  with a suitable subscript. For a set of groups which are all alike, and in the absence of interaction between the groups, this  $k$  is equal to the titration constant of Simms. The classical dissociation constants will be denoted by capital  $K$ 's.

A general treatment of some of these relations has recently been given by Wyman<sup>9</sup> as follows:

"Consider an acid in which there are in all  $m$  groups capable of dissociating a proton. Let  $A_0$  refer to the molecules from which no proton is dissociated,  $A_1$  to the molecules from which one proton is dissociated, and, more generally,  $A_h$  to the molecules from which  $h$  protons are dissociated. From a microscopic point of view, the molecules  $A_h$  are of  $\frac{m!}{h!(m-h)!}$  different kinds, which we may designate by  $A_{h1}, A_{h2} \dots$ , corresponding to the  $\frac{m!}{h!(m-h)!}$  different ways in which it is possible to choose the  $h$  groups from the total of  $m$  groups. Altogether, taking account of all values of  $h$  from 0 to  $m$ , the total number of microscopically different kinds of molecule is given by  $\sum_{h=0}^m \frac{m!}{h!(m-h)!}$ , which is the sum of the binomial coefficients for the exponent  $m$  and is equal to  $2^m$ .

"The molecules of any particular kind are in equilibrium with the completely undissociated molecules  $A_0$ , which are of course all alike. If we use parentheses to indicate concentration and denote the activity of the hydrogen ion by  $(H^+)$ , this equilibrium may be expressed in terms of the product of a series of apparent dissociation constants

$$(A_{h1}) = (A_0) \frac{k_{11} k_{21} \dots k_{h1}}{(H^+)^h} = \frac{A_0 \prod_{i=1}^h k_{hi}}{(H^+)^h} \quad (4)^*$$

Now if we add together the  $\frac{m!}{h!(m-h)!}$  expressions of this form corresponding to any given value of  $h$  we obtain an expression for the equilibrium between all the molecules  $A_h$  and the molecules  $A_0$ , namely

$$(A_h) = (A_{h1}) + (A_{h2}) + \dots = (A_0) \frac{L_h}{(H^+)^h} \quad (5)$$

The equilibrium constant  $L_h$  is defined by this equation, being equal to the sum of a series of terms, each term being a product of the form  $\prod k_{hi}$ , as given in equation 4. The number of terms in the sum is equal to the number of microscopically different kinds of molecule in the class  $A_h$ .

A dissociation constant  $K_h$  is uniquely defined by equation 5, for every value of  $h$  from 1 to  $m$ , by the relation

$$K_h = \frac{L_h}{L_{(h-1)}} = \frac{(H^+) \cdot (A_h)}{[A_{(h-1)}]} \quad (5a)$$

"If each of the  $m$  acid groups is sufficiently weak we may make the usual approximation introduced in discussions of the titration of weak acids and say that the amount of base bound by the molecules  $A_h$  is, in terms of concentration, equal to  $h(A_h)$ . The total concentration of base bound by all the molecules corresponding to all values of  $h$  then becomes

$$A_0 \sum_{h=0}^m h \frac{L_h}{(H^+)^h}$$

At the same time the total concentration of ions of all kinds is given by

$$A_0 \sum_{h=0}^m \frac{L_h}{(H^+)^h}$$

where we may write, from 5,  $L_0 = 1$ .

From these two expressions it follows that the total amount of base bound per mole of  $A$  in all its forms is given by

$$\bar{h} = \frac{\sum_{h=0}^m h L_h}{\sum_{h=0}^m L_h} \frac{1}{(H^+)^h} \quad (6)$$

If we neglect the change of activity coefficients with pH, we may treat the  $K$ 's as constants and write this expression as

$$\bar{h} = \frac{\partial \log_{10} \frac{1}{\sum_{h=0}^m \frac{L_h}{(H^+)^h}}}{\partial \log_{10} (H^+)} = \frac{\partial \log_{10} \sum_{h=0}^m \frac{L_h}{(H^+)^h}}{\partial \text{pH}} \quad (7)$$

"This expression holds whether or not the groups interact, *i.e.*, whether or not the dissociation of one group affects the dissociation of another. It is limited only by the approximations we have specifically mentioned, namely, the neglect of the change of activity coefficients with pH and the assumption regarding the base bound.

"We may consider the special case where all the  $m$  groups are characterized by the same constant  $k$  and where there is no interaction. Then  $L_h$  becomes equal to  $\frac{m!}{h!(m-h)!} k^h$ , and the expression for the base bound

This means that the amount of base bound is  $m$  times that bound by a simple monobasic acid of the same molar concentration as  $A$ .

"Consider another case, where the  $m$  groups break up into  $l$  classes containing  $n_1, n_2, n_3 \dots n_l$  members, characterized by constants  $k_1, k_2 \dots k_l$ , and where there is no interaction between any groups. Then

$$\bar{h} = n_1 \frac{\partial \log}{\partial \text{pH}} \left( 1 + \frac{k_1}{(\text{H}^+)} \right) + n_2 \frac{\partial \log}{\partial \text{pH}} \left( 1 + \frac{k_2}{(\text{H}^+)} \right) \dots + n_l \frac{\partial \log}{\partial \text{pH}} \left( 1 + \frac{k_l}{(\text{H}^+)} \right) \quad (9a)$$

$$= \frac{n_1 k_1}{(\text{H}^+) + k_1} + \frac{n_2 k_2}{(\text{H}^+) + k_2} + \dots + \frac{n_l k_l}{(\text{H}^+) + k_l} \quad (9b)$$

This may also be written as

$$\bar{h} = n_1 \alpha_1 + n_2 \alpha_2 + \dots + n_l \alpha_l \quad (9c)$$

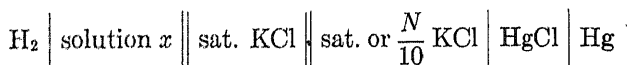
where the  $\alpha$ 's are related to the corresponding  $k$ 's according to equation 1."

It has been found experimentally that the base bound by protein molecules as a function of pH can be described by equations of the form of equations 9a, 9b, 9c; and in many cases the number of  $K$ 's required is small, being of the order of three or seven. Only when the groups of each class ionize independently of one another can the values of  $k_1, k_2 \dots k_l$ , in equation 9, be regarded as equal to the intrinsic acidity constants of individual acidic groups. The problem of the interaction of the groups further complicates the theory and is considered in connection with the effect of ionic strength on protein titration curves (p. 468).

#### Calculation of Free and Bound Acid and Base

Throughout the range from approximately pH 4 to 10, the concentration of free hydrogen and hydroxyl ions in the system is negligibly small in comparison with the amount bound by the protein; that is, unless the protein concentration is very small. In strongly acid solutions the concentration of free  $\text{H}^+$  ions and in strongly alkaline solutions the concentration of  $\text{OH}^-$  ions is, however, significant in comparison with the amount which has reacted with the protein.

Measurements of pH on the hydrogen electrode are made with the cell.



EMF measurements made on such a cell determine hydrogen ion activities,\* which we have denoted by  $(\text{H}^+)$ .

An activity coefficient for the hydrogen ion must then be evaluated in order to determine the hydrogen ion concentration. This may be done for a protein solution by comparison with a solution of a pure strong acid or base containing no protein, since there is abundant evidence for such pure electrolytes that in dilute aqueous solution their activity coefficients are adequately described by the interionic force theory of Debye and Hückel, assuming that the electrolyte is completely dissociated. This assumption is supported by the experimental measurements, not only in water, but in methyl and ethyl alcohol<sup>10</sup>. Thus, if solution *x* in the above cell is HCl at concentration  $C_1$ , the hydrogen ion concentration is  $C_1$  moles per liter. If the hydrogen ion activity obtained from the EMF of this cell is  $(H^+)_1$ , then the activity coefficient of the hydrogen ion is

$$\gamma_H = \frac{(H^+)_1}{C_1}.$$

If now solution *x* be composed of HCl at concentration  $C_1$ , plus a given amount of protein, a new value of hydrogen ion activity,  $(H^+)_2$ , is obtained. It has been found that simple and consistent results are obtained in protein titration curves if it be assumed as a first approximation that  $\gamma_H$  in the second solution is the same as in the first; in other words, that

$$\frac{(H^+)_1}{C_1} = \frac{(H^+)_2}{C_2}; \quad \text{or} \quad \log \frac{C_2}{C_1} = \text{pH}_1 - \text{pH}_2 \quad (10)$$

from which the concentration of hydrogen ions in the second solution is immediately determined. The difference,  $C_1 - C_2$ , gives the moles of hydrogen ion bound by the protein per liter of solution. Although the results so obtained are generally very self-consistent, this of course does not prove that they are correct. This procedure is, however, as free from objection as certain alternatives which have been proposed.

It is interesting to note that essentially this procedure was adopted in 1898 by Bugarszky and Liebermann<sup>11</sup>, who were the first investigators to demonstrate by EMF studies that proteins combine with acids and bases. They measured the electromotive force between hydrogen electrodes, one immersed in a cell containing 0.05*N* hydrochloric acid, the other in a cell containing the same concentration of acid and varying amounts of egg albumin. From the initial concentration of the acid and the electromotive force measurements, the concentrations of free acid were calculated by means of the relation:

$$\text{EMF} = \frac{RT}{nF} \ln \frac{C_2}{C_1}$$

The difference between the free acid and the total acid present yields the acid combined with the protein. A recalculation of Bugarszky and Liebermann's results is tabulated below, in part because of historical significance and in part because the values obtained are only slightly lower than those obtained later by the most careful investigators.

A similar treatment may be applied to the calculation of the bound hydroxyl ions in alkaline solution. If we represent the isoelectric protein dipolar ion as  $^+HP^-$ , its reaction with hydroxyl ion may be written  $^+HP^- + OH^- \rightarrow P^- + H_2O$ , and the "bound hydroxyl ions" represent the number of protons removed from the isoelectric protein by this type of reaction.

In a solution of pure  $Na^+ OH^-$  or  $K^+ OH^-$  of concentration  $C_1$ , the concentration of  $OH^-$  ions is likewise  $C_1$ . The pH measured on the hydro-

TABLE 2. ACID COMBINING CAPACITY OF EGG ALBUMIN CALCULATED FROM THE RESULTS OF BUGARSKY AND LIEBERMANN\*

Albumin in system $\frac{p}{\text{grams}}$	HCl in system $\frac{a}{N}$	HCl free $\frac{b}{N}$	HCl combined $\frac{a-b}{N}$	HCl combined by 10 grams albumin $\frac{(a-b)/p}{N}$
6.4	0.05	0.0027	0.0473	0.0074
3.2	0.05	0.0262	0.0238	0.0074
1.6	0.05	0.0373	0.0127	0.0079
0.8	0.05	0.0436	0.0064	0.0080
0.4	0.05	0.0468	0.0032	0.0080

From Cohn, E. J., *Physiol. Rev.*, 5, Table 4, p. 384 (1925).

\* Bugarszky, S., and Liebermann, L., *Pflüger's Archiv*, 72, 51 (1898).

gen electrode yields the H ion activity  $(H^+)_1$ . The  $OH^-$  ion activity,  $(a_{OH})_1 = Kw/(H^+)_1$  and the activity coefficient  $\gamma_{OH}$  is

$$(\gamma_{OH})_1 = \frac{(a_{OH})_1}{(c_{OH})_1} = \frac{K_w}{(H^+)_1 C_1}$$

Now consider solution 2, containing the same concentration of strong alkali, and also a given amount of protein. The total concentration of alkali is again  $C_1$ ; the concentration of free  $OH^-$  ions, to be determined, is denoted as  $C_2$ . The measured pH of this solution is taken as the H ion activity,  $(H^+)_2$ . We now make the postulate, similar to that made in the acid solution discussed above, that

$$(\gamma_{OH})_1 = (\gamma_{OH})_2 = \frac{K_w}{(H^+)_1 C_1} = \frac{K_w}{(H^+)_2 C_2}$$

Hence  $(H^+)_1 C_1 = (H^+)_2 C_2$ ; or



This treatment eliminates any need of employing the dissociation constant of water,  $K_w$ , in the calculations. The only quantities which must be known are the analytically determined values of the concentration of alkali in solutions 1 and 2, and of protein in solution 2; and the pH values of the two solutions. This is essentially the procedure adopted by Cohn and Berggren<sup>12</sup> in 1923, in their studies on casein.

This analysis can be criticized on theoretical grounds. The activity coefficient of any ion is a function of the nature and concentration of all the ions present. In the above treatment, in the acid solution the concentration of chloride anions is the same in solutions 1 and 2. The equivalent concentration of cations is, of course, also the same in both; but in solution 1 the cations are all monovalent  $H^+$  ions, while in solution 2 they are largely polyvalent protein ions. The total ionic strength in solution 2 should thus be greater than in solution 1, and the activity coefficient,  $\gamma_H$ , might be expected to be smaller. The same criticism holds for the interpretation of the measurements made in alkaline solutions. This point was carefully considered by Cohn and Berggren<sup>12</sup> (see especially Table 3 of their paper). They did not, however, suggest any departure in practice from the procedure here adopted.

There is indeed at present no simple method of evaluating the contribution to the ionic strength of large polyvalent ions with widely separated charges. The wide separation of the charges should certainly cause the protein ion to affect the ionic strength of the solution much less than if the same net charge were concentrated within an ion of very small radius<sup>13</sup>. The justification of equations 10 and 11 is essentially pragmatic; they give consistent and regular results for the maximum acid binding by proteins, even when the concentration of protein and of total acid are independently varied over a considerable range. Further advance of knowledge may well lead to an equation more accurate than equation 1, but it is not likely that any drastic revision of our present views on the binding of acids by proteins will be required.

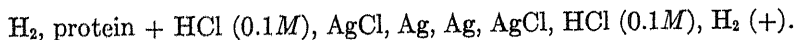
The effect of any errors in equations 10 and 11 can be minimized by making the unknown concentration  $C_2$  as small as possible, relative to the known concentration  $C_1$ ; that is, by making the concentration of free acid small compared to the acid bound by the protein. This can best be achieved by making the protein concentration as high as possible when working in strongly acid (or alkaline) solutions. This of course increases the uncertainty arising from the effect of the protein on the activity coefficient of the hydrogen ion; but this disadvantage is much more than

counterbalanced by the advantage of reducing  $(H^+)_2$  and  $C_2$  to values small compared to  $(H^+)_1$  and  $C_1$ .

Another device which is often of value in eliminating uncertainties regarding activity coefficients is to conduct the titration in the presence of a large excess of a neutral salt, such as sodium or potassium chloride, at constant ionic strength, and to carry out EMF determinations on pure strong acid (or alkali) in the presence of the same salt at the same ionic strength. Equations 10 and 11 should be much more rigorously applicable in such systems than in the absence of added salt, since the added salt, at high ionic strength, becomes the major factor in determining the activity coefficients of all ions present, and also the liquid junction potential. "It has been shown that the Nernst formula for electromotive force, the solubility product and mass action law in the case of a complicated ionic equilibrium are applicable in their classical form to such concentrated salt solutions, the reason for this simplicity being the practical constancy of the activity coefficients in the practically constant medium. Utilization of these results would mean in many cases a great simplification in problems pertaining to electrolytic solutions" (reference 14, p. 431). Cannan<sup>15</sup> has applied this method to egg albumin and lactoglobulin in the course of extensive studies of the effect of ionic strength on protein titration curves, which are discussed in more detail later in this chapter. Likewise, Cohn, Green and Blanchard<sup>16</sup> studied the titration curve of hemoglobin both in the absence of salt and in the presence of molar sodium chloride over a wide range of pH. A practical limit, however, was set to this method of titrating in acid solution by the fact that the hydrochloride of hemoglobin is insoluble in the presence of excess chloride ion. The same difficulty arises with many other proteins—for instance, edestin and myosin—and, therefore, involves a serious practical limitation on this method which is theoretically so advantageous. Such difficulties, however, seldom arise for proteins in neutral or alkaline solutions and this method is undoubtedly adapted to much wider application than has been made of it in the past.

A very careful and critical study of the maximum combining capacity of proteins with acid has been made by Hitchcock<sup>17</sup> based on the electromotive force of cells without liquid junction involving hydrogen electrodes and silver-silver chloride electrodes. We quote from his discussion.

"These cells were essentially of the following type:



<sup>14</sup> Brönsted, J. N., *Trans. Faraday Soc.*, 23, 416 (1927). See also Harned, H. S., "*The Electrochemistry of Solutions*" in *Methods in Biological Chemistry*, 2nd edition, D. van Nostrand Co., New York, Vol.

Cells of this type, without protein, had been used with marked success by Lewis, Harned and others in studying the activity coefficients of single or mixed inorganic strong electrolytes. This double cell is a concentration cell without transference, and its electromotive force is given by the exact thermodynamic equation,

$$E = k \log \frac{m^2 \gamma_0^2}{m_H m_{Cl} \gamma^2} \quad (12)$$

Here  $k$  is  $2.3026 RT/F$ ,  $m$  and  $\gamma_0$  are the molality and the mean activity coefficient of the ions of the acid without protein,  $m_H$  and  $m_{Cl}$  are the molalities of the hydrogen and chloride ions in the protein solution, and  $\gamma$  is the mean activity coefficient of these ions in the protein solution. If the protein concentration is expressed as  $g$  grams of protein per kilogram of water, and if one gram of protein binds  $x$  equivalents of hydrogen ions and  $y$  equivalents of chloride ions, the product  $m_H m_{Cl}$  may be replaced by  $(m - gx)(m - gy)$ , or

$$E = k \log \frac{m^2 \gamma_0^2}{(m - gx)(m - gy) \gamma^2} \quad (13)$$

"In order to calculate values for  $x$  by means of equation 13, it is necessary to make some assumption about the values of  $y$  and  $\gamma$ . Three sets of assumptions were tried:

I.  $y = 0. \quad \gamma = \gamma_0$

$$E = k \log \frac{m}{(m - gx)} \quad (14)$$

II.  $y = \text{constant.} \quad \gamma = \gamma_0$

$$E = k \log \frac{m^2}{(m - gx)(m - gy)} \quad (15)$$

III.  $y = 0. \quad \log \gamma = \log \gamma_0 - bg.$

$$E = k \log \frac{m}{m - gx} + 2kbq. \quad (16)$$

"The results of these measurements and calculations for three proteins are given in Table 3. The assumption of constant activity coefficients is similar to that used in getting protein dissociation curves from pH data. As applied to the data obtained without liquid junction, this assumption does not give constant combining capacities for hydrogen ions (I) unless

law of Harned<sup>18</sup> and Güntelberg<sup>19</sup>, which applies to mixtures such as those of hydrochloric acid with sodium chloride, at constant total molality. A similar rule was found by Failey<sup>20</sup> to represent the effect of edestin in acid solutions on the activity coefficients of thallos chloride.

"The second and third assumptions give the values of  $x$  which are constant and practically identical. This agreement was puzzling until it was realized that equation II can be reduced to the same form as III by expanding the logarithm of one factor into a series and neglecting higher terms. Hence the data do not indicate whether II or III is more likely to be correct; the analogy with the linear law just mentioned seems to favor III. For gelatin and edestin, the values of  $x$  are in excellent agreement

TABLE 3. ACID COMBINING CAPACITIES OF PROTEINS FROM THE ELECTROMOTIVE FORCE OF CELLS WITHOUT LIQUID JUNCTION AT 30°C.\*

Protein	$g$ = protein per kg H <sub>2</sub> O	$m$ = HCl per kg H <sub>2</sub> O	$E$ = EMF of double cell	$10^3 x = H^+$ bound per gm protein (milliequivalents)		
				I	II	III
Casein	<i>gm</i>	<i>moles</i>	<i>millivolts</i>			
	20.0	0.1000	6.6	1.116	0.814	0.800
	30.0	0.0999	10.2	1.076	0.801	0.796
	40.0	0.0999	14.1	1.042	0.796	0.795
	50.0	0.1000	18.5	1.015	0.799	0.802
					( $10^3\gamma = 0.36$ )	( $10^3b = 8.5$ )
Edestin	18.55	0.1020	9.0	1.602	1.336	1.326
	42.35	0.1050	24.4	1.505	1.345	1.340
	62.7	0.1055	47.6	1.411	1.339	1.340
					( $10^3\gamma = 0.35$ )	( $10^3b = 8.0$ )
Gelatin	50.3	0.1000	19.5	1.046	0.957	0.958
	70.4	0.1000	32.8	1.016	0.961	0.961
	90.5	0.1000	57.1	0.981	0.958	0.959
					( $10^3\gamma = 0.17$ )	( $10^3b = 3.85$ )

\* From Hitchcock, D. I., *Cold Spring Harbor Symp. Quant. Biol.*, 6, Table I, p. 28 (1938).

\* Hitchcock, D. I., *J. Gen. Physiol.*, 16, 357 (1932-33), and Hitchcock, D. I., in Schmidt, C. L. A., "The Chemistry of the Amino Acids and Proteins," Springfield, Illinois (1938).

with those previously obtained by other methods; it seems likely, therefore, that the combining capacities for hydrogen ions are correct, although the question whether chloride ions combine with proteins is still undecided" (reference 17, p. 27-28).

In Table 4 of Chapter 15 we have summarized the data obtained on cells with liquid junction for these same proteins and compared them with the results obtained by Hitchcock. This table also gives comparative figures for the analytical data on the histidine, arginine and lysine content of all these proteins in comparison with the acid binding capacity, as

determined from EMF measurements. Similar data are also given for zein, gliadin, insulin, hemoglobin and limulus hemocyanin and other proteins, either because of their general importance or because of the accuracy of the analytical data upon them, although they have not been studied in cells without liquid junction. In general, the agreement between values obtained with and without liquid junction may be considered very satisfactory, considering the present state of our knowledge. The discrepancy with regard to casein depends entirely upon the divergence in the values obtained by different investigators from measurements in cells with liquid junction. Hitchcock's determinations on cells without liquid junction fall well within the limits of this divergence and correspond very well with the total content of dibasic amino acids, as determined by isolation procedures.

Another uncertainty in the interpretation of EMF measurements has been pointed out by several authors. It is always possible, when work is carried out at a pH as acid as 1 or 2, or as alkaline as 12 or 13, that some hydrolysis of peptide or other linkages may take place, increasing the maximum capacity for binding acid or base above that of the native protein. Synthetic peptides, in solution at room temperature, show no appreciable hydrolysis within short periods of time; but it is always possible that more labile peptide or other linkages exist in the protein molecule. In any case, such linkages must be few, since the maximum acid combining capacity of many proteins, as will be seen from the later discussion in this chapter, is little if at all greater than their known content of dibasic amino acids. Such discrepancies as do exist are quite as likely to be due to inadequate analytical estimates as to partial hydrolysis of the protein during titration.

It is important, however, in determining acid and base combining capacity from electromotive force measurements, to determine whether there is any change of EMF with time, when a given solution is observed over a period of several hours or even for one or two days.

#### **Equilibria between Different Ionic Forms in Polyvalent Acids and Ampholytes**

The interpretation of protein titration curves involves the analysis of several factors which are not present or are unimportant in studies on amino acids. Protein molecules contain only a few different kinds of ionizing groups, but they generally contain many different groups of one or more kinds. Therefore, in a given solution, many different ionic forms of the proteins may exist in equilibrium with one another. Some aspects of the problems which thus arise have already been discussed in Chapter

groups in a protein molecule is given by equation 9, as a function of the pH, the titration constants of the protein, and the number of groups in the molecule corresponding to each titration constant.  $\bar{n}$  is thus the base bound by the protein, taking the point of maximum acid binding as the zero of reference. If  $n$  is the positive charge on the protein when the maximum amount of acid is bound, then the mean net charge on the protein at any pH is  $n - \bar{n}$ . Corresponding to this mean value, however, there must be in general a large number of different ionic forms of the protein in equilibrium with one another. Thus (see Wyman's discussion, p. 451) if the protein contains altogether  $m + n$  groups capable of yielding up a proton, the number of possible different configurations from which  $h$  protons have been removed is  $\frac{(m+n)!}{h!(m+n-h)!}$ . Thus for horse CO-

hemoglobin in which  $m + n$  may be taken as 174 (probably a slight underestimate), the molecule is isoelectric when 99 protons have been removed (p. 478). Thus the number of possible isoelectric forms is approximately  $\frac{174!}{99!75!} = 6 \times 10^{49}$ . Most of the forms corresponding to this vast number must obviously be of negligible importance. All the isoelectric forms present to any appreciable extent must carry negatively charged carboxyl groups, uncharged phenolic hydroxyl and sulfhydryl groups, and positively charged ammonium and guanidinium groups. The actual state of the molecule in the solution depends chiefly on the distribution of charge among the 33 imidazole residues of histidine which are present (p. 480). Hemoglobin is isoelectric when 12 of these 33 residues have lost protons. The number of different configurations which correspond to this state of total charge is

$$\frac{33!}{12! 21!} = 354,817,320$$

and any one of these possible forms has a reasonable chance of existing in the solution at any moment. Where the actual configuration of charges in a protein molecule is important we must reckon, not with a single isoelectric form, but with a large number of different forms.

All these forms of a given net charge are in equilibrium, not only with each other, but with other forms of greater and less net charge. In any isoelectric protein solution, for instance, protein anions and cations are present in considerable amounts. It is of interest in many connections—for example, in considering the conductivity of protein solutions or in

\* interpreting solubility experiments on proteins—to be able to evaluate

1. **The Equation of Linderstrøm-Lang.** That a problem of this sort existed was apparently first clearly stated by Linderstrøm-Lang<sup>21</sup>. He pointed out that in a polyvalent acid or ampholyte, as the number of acid or basic groups increases, "so also an increase takes place in the number of dissociation stages, and since the dissociation constants lie close together, there will be several different types of ions with different valencies determining the mean valency" (reference 21, p. 22). He proceeded to develop an electrostatic theory of the effect of ionic strength on the titration constants of polyvalent acids and ampholytes, which is considered on p. 472.

Recently Linderstrøm-Lang has given a quite general equation which gives a simple measure of the deviation of individual ions in a solution from the average state of the ions in the system.\* He has derived this from a treatment essentially identical with that already considered (p. 451). His equation may be immediately derived by differentiating the general equation for the average number of protons  $\bar{h}$ , as given by equation 6, with respect to pH.

$$\begin{aligned} \frac{\partial \bar{h}}{\partial \text{pH}} &= - \frac{\partial(n - \bar{h})}{\partial \text{pH}} = -2.303 \frac{\partial \bar{h}}{\partial \ln (\text{H}^+)} = -2.303 (\text{H}^+) \frac{\partial \bar{h}}{\partial (\text{H}^+)} \\ &= 2.303 \frac{\left[ \sum \frac{K_h}{(\text{H}^+)^h} \cdot \sum \frac{h^2 K_h}{(\text{H}^+)^h} - \left( \sum \frac{h K_h}{(\text{H}^+)^h} \right)^2 \right]}{\left( \sum \frac{K_h}{(\text{H}^+)^h} \right)^2} = 2.303 [\bar{h}^2 - (\bar{h})^2] \end{aligned} \quad (18)$$

The difference between the mean square value  $\bar{h}^2$ , and the square of the mean value,  $(\bar{h})^2$ , is the standard deviation† of the function  $h$ . It is the simplest index of the proportion of molecules whose net charge at any moment differs from the mean net charge of all the molecules in the solution.

At pH values for which the concentration of free hydrogen or hydroxyl ions is negligibly small, the function  $\frac{\partial \bar{h}}{\partial \text{pH}}$  is equal to the "buffer value" of Van Slyke<sup>22</sup>, per mole of protein.

We may now employ the value of  $\bar{h}$ , given by equations 9a, 9b and 9c. These equations may be compactly written as equation 9:

$$\bar{h} = \sum_i \frac{n_i k_i}{(\text{H}^+) + k_i} = \sum_i n_i \alpha_i. \quad (9)$$

<sup>21</sup> Linderstrøm-Lang, K., *Comment. chem. trav. lab. Carlsberg*, 15, No. 7 (1924).

<sup>22</sup> Several years ago, by Dr. Linderstrøm-Lang, but has not previously which these problems were discussed, at the Department of Physical Chemistry, University of Copenhagen, in 1939. Linderstrøm-Lang's treatment is also given in a somewhat

Then, following Van Slyke's treatment, the buffer value is found to be

$$\frac{\partial \bar{h}}{\partial \text{pH}} = 2.303 \sum_i \frac{n_i k_i (\text{H}^+)}{(k_i + (\text{H}^+))^2} = 2.303 \sum_i n_i \alpha_i (1 - \alpha_i) \quad (19)$$

and, combining 18 and 19

$$\bar{h}^2 - (\bar{h})^2 = \sum_i n_i \alpha_i (1 - \alpha_i) \quad (20)$$

We may apply this equation, for instance, to hemoglobin at its isoelectric point, which was at approximately pH 6.4 in the experiments of Cohn, Green and Blanchard<sup>16, \*</sup>. Two sets of histidine groups, corresponding to the titration constants denoted by them as  $\text{pK}'_4 = 5.7$  (containing 13 groups) and  $\text{pK}'_5 = 7.5$  (containing 20 groups) are by far the most important in the application of equation 20. For the former set,  $\alpha = 0.833$  at pH 6.4, for the latter,  $\alpha = 0.074$ . Hence

$$\bar{h}^2 - (\bar{h})^2 = 13(.833 \times .167) + 20(.74 \times .926) = 3.18$$

A small additional contribution is made by the 87 carboxyl groups, for which a mean value of  $1 - \alpha$  between .002 and .004 may be taken ( $\text{pK}'$  between 3.7 and 4.0), giving an additional contribution of 0.17 to 0.34. The contributions from groups having  $\text{pK}'$  values of 10.8 or greater are negligible. Thus  $\bar{h}^2 - (\bar{h})^2 = 3.4$  to 3.5; and the root mean square value of the net charge per molecule of hemoglobin at the isoelectric point is approximately 1.85.

This treatment is perfectly general, provided that the  $k$  values determined for the protein are not a function of pH; this requirement is essentially fulfilled, provided the titration from which they are derived is conducted in a medium of constant temperature, chemical composition and ionic strength (except for the small amount of acid and base added).

We may now consider a second approach to this problem, which is more restricted in its application, but when applicable gives a more detailed picture of the state of the system.

**2. The Probability Distribution Function for a System of Independently Ionizing Groups.** In this treatment, we confine ourselves to systems which fulfil the conditions assumed in deriving equations 9a, 9b and 9c; that is, the groups in any one class, corresponding to one  $\text{pK}'$  value, do not interact. The quantity  $\alpha_k$  (equations 1 and 9c) for the  $k$ 'th class of groups, denotes at any pH the fraction of the total number of groups in this class from which a proton has been removed. But  $\alpha_k$  also denotes the probability that any individual group in this class, chosen at random, would be found to have lost a proton. If the number of groups in this class per



molecule is  $n_k$ , then the probability of a particular configuration, in which a specified set of  $q_k$  individual groups have lost their protons, while the remaining  $n_k - q_k$  have not, is

$$p(q_k) = \alpha_k^{q_k} (1 - \alpha_k)^{n_k - q_k} \quad (21)$$

The number of ways in which  $q_k$  protons can be removed from the  $n_k$  groups is  $\frac{n_k!}{q_k!(n_k - q_k)!}$ , and by our initial postulate all of these possible ways are equally probable. Thus the fraction of all the molecules in the system which have lost  $q_k$  protons of the  $k$ 'th class, that is, the total probability of the state  $q_k$  is:

$$\begin{aligned} P(q_k) &= \frac{n_k!}{q_k!(n_k - q_k)!} \alpha_k^{q_k} (1 - \alpha_k)^{n_k - q_k} \\ &= \frac{n_k!}{q_k!(n_k - q_k)!} \left( \frac{\alpha_k}{1 - \alpha_k} \right)^{q_k} (1 - \alpha_k)^{n_k} \end{aligned} \quad (22)$$

where  $q_k$  may assume the values 0, 1, 2  $\dots$   $n_k$ . Thus we need only evaluate the function  $P(q_k)$  for all values of  $q_k$  from 0 to  $n_k$ , to determine the relative numbers of molecules in the different states of net charge.\*

Similarly if we consider the set of  $n_j$  residues corresponding to the  $j$ 'th titration constant:

$$P(q_j) = \frac{n_j!}{q_j!(n_j - q_j)!} \left( \frac{\alpha_j}{1 - \alpha_j} \right)^{q_j} (1 - \alpha_j)^{n_j} \quad (23)$$

Thus the probability of the state in which  $q_k$  protons have been removed from the  $k$ 'th class, and  $q_j$  from the  $j$ 'th class is

$$P(q_j, q_k) = P(q_k) \cdot P(q_j) \quad (24)$$

**A Simple Case.** Consider a hypothetical peptide, composed of five glutamyl residues ( $pK' = 4.5$ ) and three lysyl residues ( $pK' = 10.5$ ) all widely separated from each other by monoamino-monocarboxylic acids, so that no one of the charged groups appreciably affects the ionization of the others. (The terminal  $\alpha$ -amino and carboxyl groups are tied off so that they do not react with acids or bases.) This molecule will be isoelectric when  $\bar{h} = 3$ , that is, when on the average three-fifths of the carboxyl groups in the system have lost their protons. At the isoelectric point ( $pH\ 4.676$ )  $\alpha = 0.6$ , and  $\alpha/(1 - \alpha) = 1.5$ . At this  $pH$ , all the lysyl

\* If  $P(q)$  is known for any one value of  $q$  it may readily be calculated for other values by means of the relation:

residues in the system may be assumed to be positively charged; the error in this assumption is negligible. Then, utilizing equation 22, the state of the system is given by

$$P(q) = \frac{5!}{q!(5-q)!} (1.5)^q (0.4)^5; \quad q = 0, 1, 2, 3, 4, 5,$$

where  $q$  is the number of carboxyl groups which have lost protons in any given molecule. When carried through, the calculation yields Table 4.

Thus at any moment only about one-third of all the molecules present at the isoelectric point are truly isoelectric; the remaining two-thirds carry a net charge. The mean square value of this net charge is found from equation 20 to be 1.2; the same value may be calculated from Table 4 by multiplying each value of  $P(q)$  by the square of the corresponding net charge, and summing all the terms.

TABLE 4. DISTRIBUTION OF CHARGED FORMS IN A PEPTIDE CONTAINING FIVE CARBOXYL GROUPS AND THREE POSITIVE CHARGES, AT THE ISOELECTRIC POINT

$q$ = no. of ionized COO- groups in molecule	Net charge $Z = 3 - q$	$P(q)$ = fraction of molecule in state $q$
0	+3	.0102
1	+2	.0768
2	+1	.231
3	0	.346
4	-1	.259
5	-2	.0775

Such calculations are not, of course, confined to the isoelectric point; they can be carried out equally well at any other pH value, and any state of mean net charge.

#### Equilibria between Different Charged Forms in Isoelectric Hemoglobin

In hemoglobin at pH 6.4, the value of  $\bar{n}$  is 99, and the protein is isoelectric under the conditions of the titration shown in Fig. 1, (p. 466). All the 87 carboxyl groups found will be assumed to carry negative charges at this point. Thus the molecule is isoelectric when on the average 12 of the 33 imidazole groups on the histidine residues in each molecule have lost protons. These residues are of two classes: Class A (13 residues) with  $pK'$  5.7,  $\alpha_A = .833$ , and Class B (20 residues) with  $pK'$  7.5,  $\alpha_B = 0.74$ . Then, if we denote the fraction of hemoglobin molecules in which five residues of Class A have lost protons by  $P_A(5)$ , etc., and similarly for Class B, we obtain, from equation 23, Table 5. The proportion of other states of charge in Classes A and B is negligible at this pH.

The fraction of all the hemoglobin molecules in which a total of  $q$  groups

by  $P_T(q)$ . This function may be evaluated from equation 24, and the data of Table 5. Thus, for example, if  $q = 12$ , we have:

$$P_T(12) = P_A(0) \cdot P_B(12) + P_A(1) \cdot P_B(11) \\ + \cdots P_A(11) \cdot P_B(1) + P_A(12) \cdot P_B(0)$$

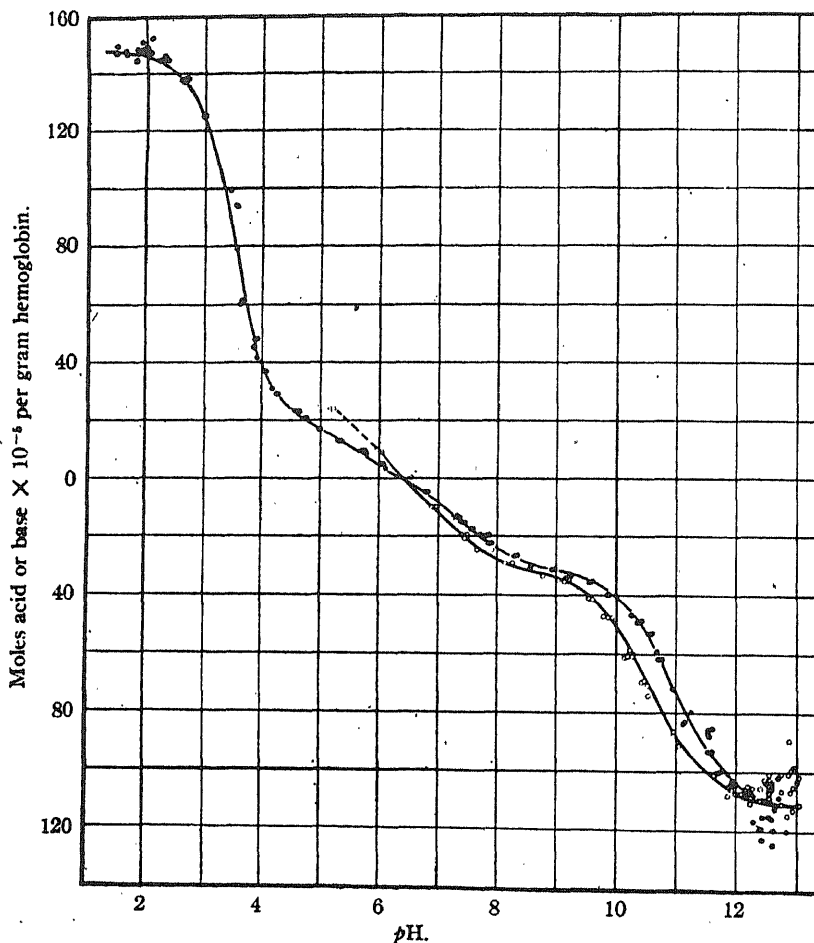


Figure 1. The titration curve of carboxyhemoglobin of the horse in the absence of added salt (solid circles) and in the presence of 1 *M* NaCl (open circles).

From Cohn, E. J., Green, A. A., and Blanchard, M. H., *J. Am. Chem. Soc.*, **59**, 509 (1937).

Many of these terms (for instance, the first six in the above series) are so

Thus it appears that less than a quarter of the hemoglobin molecules, in an isoelectric solution, are actually isoelectric at any moment. The exact values given in Table 6 depend on the specific assumptions regarding  $pK'$  values and the number of groups involved; but any treatment compatible with the general form of the titration curve near the isoelectric point would lead to a closely similar result, even if it were assumed that electrostatic interaction between the ionizing groups is considerable. Here

TABLE 5. DISTRIBUTION OF CHARGE AMONG HISTIDINE RESIDUES OF HEMOGLOBIN OF CLASS A ( $pK'$  5.7) AND CLASS B ( $pK'$  7.5) AT pH 6.4

Class A		Class B	
$P_A$	(5) = .00016	$P_B$	(0) = .215
$P_A$	(6) = .00107	$P_B$	(1) = .344
$P_A$	(7) = .0103	$P_B$	(2) = .262
$P_A$	(8) = .0385	$P_B$	(3) = .1255
$P_A$	(9) = .1071	$P_B$	(4) = .0426
$P_A$	(10) = .214	$P_B$	(5) = .0109
$P_A$	(11) = .292	$P_B$	(6) = .0022
$P_A$	(12) = .243	$P_B$	(7) = .00035
$P_A$	(13) = .0935		

TABLE 6. DISTRIBUTION AMONG FORMS OF DIFFERENT NET CHARGE IN HEMOGLOBIN AT pH 6.4

$q$ = no. of uncharged imidazole residues per molecule	$Z$ = net charge = $12 - q$	Fraction of molecules in state $q$ = $P_T(q)$	Contribution to total mean square net charge = $Z^2 P_T(Z)$
6	+6	.0003	.0108
7	+5	.0026	.0650
8	+4	.0126	.2016
9	+3	.0390	.3510
10	+2	.0942	.3768
11	+1	.1698	.1698
12	0	.2240	.0000
13	-1	.2120	.2120
14	-2	.1430	.5720
15	-3	.0699	.6291
16	-4	.0265	.4240
17	-5	.0071	.1775
18	-6	.0018	.0648
19	-7	.0003	.0147
Total.....		1.0041	3.190

The sum of all values of  $P_T(q)$  should be exactly 1; the value obtained (1.0041) shows the slight deviation from this value due to approximations made in the numerical calculations.

again, for a protein as for a peptide, the method of analysis adopted is not restricted to the isoelectric point, but is equally applicable to any state of charge on the molecules.

The existence of such equilibria between the different charged forms is of importance in several respects. It should lead to higher values for the

experiments, account should be taken of the fact that at any pH many different ionic forms are present in solution. The activity coefficients of these different forms may be very differently affected by such factors as the ionic strength and the dielectric constant; while the measured solubility at any point represents the summation of the *concentration* of all these forms under the given conditions. The principles involved here have already been set forth in Chapter 3 (see especially equations 187-197 inclusive).

On the other hand, it is not to be expected that electrophoresis measurements on a chemically homogeneous protein preparation under given conditions of temperature, pH and ionic strength, should reveal any non-uniformity of mobility among the protein molecules, due to this charge distribution. Any individual molecule in the system is constantly giving up and taking on protons, so that the time average of its net charge, considered over an appreciable time interval in which many proton exchanges take place, is identical with the mean net charge of all the molecules in the system. The influence of the non-uniform distribution of charge among the molecules should thus affect their random Brownian motion when they are placed in an electric field. It should not alter the *mean* electrokinetic potential,  $\psi_0$ , (Chapter 25) which determines the electrophoretic mobility.\*

The existence of this charge distribution function may have significant implications for the study of protein solutions in alternating fields of high frequency (see Chapter 22). The rotation of the polar protein molecules in the field leads to absorption of energy by the dielectric. If ions carrying a net charge are present, in equilibrium with the isoelectric protein, there should be additional energy absorption, due to the translational motion of these ions in the alternating field, which is opposed by the frictional resistance of the medium. This translation, however, can occur only if the time of alternation of the field is short compared with the time required for proton exchange between the protein and the surrounding medium. Otherwise we must deal, as in the case of electrophoresis in a steady field (discussed above), with the time average of the net charge on a protein molecule, which vanishes at the isoelectric point. As yet there is no reliable knowledge of the rates of proton exchange in solutions, and further analysis of this problem must remain for the future.

### The Effect of Ionic Strength on Protein Titration Curves

It has long been known that the form of the titration curve of any given protein is markedly changed by variations in the ionic strength of the

TABLE 7. DISSOCIATION CURVES OF EGG ALBUMIN AT CONSTANT  $I/2$  (KCl)

$\bar{Z}$  = equi.  $[H^+]$  bound per 45,000 gm. Egg Albumin

Ionic strength $I/2$	0.0085	0.017	0.033	0.067	0.133	0.267	0.67	1.19	2.38
$\bar{Z} = n - k$	pH	pH	pH	pH	pH	pH	pH	pH	pH
39						2.03	2.40	2.43	2.53
38				2.00	2.17	2.31	2.58	2.68	2.80
36			2.00	2.28	2.48	2.63	2.88	2.95	3.04
34			2.24	2.50	2.70	2.83	3.09	3.16	3.24
32			2.46	2.70	2.88	3.02	3.24	3.31	3.40
30			2.65	2.86	3.04	3.18	3.38	3.45	3.54
28			2.81	3.03	3.18	3.32	3.50	3.57	3.66
26			2.96	3.17	3.32	3.43	3.61	3.68	3.77
24			3.11	3.31	3.45	3.55	3.72	3.79	3.87
22			3.26	3.44	3.57	3.67	3.82	3.89	3.97
20	3.13	3.29	3.41	3.57	3.69	3.78	3.92	3.99	4.07
15	3.55	3.66	3.75	3.88	3.98	4.05	4.17	4.23	4.32
10	3.90	4.06	4.12	4.21	4.28	4.33	4.42	4.48	4.56
8	4.14	4.22	4.26	4.34	4.40	4.44	4.52	4.58	4.66
6	4.32	4.38	4.41	4.48	4.53	4.56	4.63	4.69	4.76
4	4.50	4.55	4.57	4.62	4.65	4.69	4.75	4.79	4.87
2	4.69	4.72	4.74	4.77	4.79	4.83	4.87	4.92	4.99
0	4.90	4.91	4.92	4.93	4.95	4.97	5.00	5.05	5.13
-2	5.14	5.14	5.14	5.12	5.14	5.15	5.17	5.21	5.29
-4	5.39	5.38	5.36	5.33	5.33	5.35	5.35	5.39	5.47
-6	5.68	5.65	5.62	5.59	5.56	5.57	5.58	5.62	5.70
-8	6.05	6.00	5.96	5.91	5.89	5.87	5.86	5.90	5.98
-10	6.54	6.45	6.39	6.33	6.29	7.28	6.26		6.39
-12	7.12	6.90	6.97	6.91	6.87	6.82	6.80		6.93
-14			(8.1)	(8.0)	(7.9)	(7.9)	(7.8)		(8.0)
-16			(9.4)	(9.2)	(9.1)	(9.1)	(9.0)		(9.2)
-18			9.94	9.80	9.73	9.69	9.62		9.75
-20			10.29	10.15	10.06	10.02	9.93		
-22			10.56	10.41	10.31	10.26	10.15		
-24			10.79	10.63	10.53	10.47	10.23		
-26			11.02	10.84	10.73	10.65	10.52		
-28			11.22	11.04	10.92	10.83	10.69		
-30			11.42	11.23	11.12	11.02	10.88		
$\alpha_s$	-0.0205	-0.014	-0.007	0.000	0.005	0.009	0.014	0.0155	0.016
$\beta_0(KCl)$	0.00	0.01	0.02	0.03	0.04	0.07	0.10	0.14	0.22
$\beta_0(CaCl_2)$			-0.05	-0.06	-0.08	-0.10	-0.14		
$(\frac{\Delta pH}{\Delta \bar{Z}})_{pI}$	.085	.078	.071	.064	.059	.055	.051	.050	.049
$w_{exp.}$	.0585	.0505	.0425	.0345	.0285	.024	.019	.0185	.0175
$w_1$	.0696	.0592	.0484	.0385	.0299	.0227	.0153	.0118	.0086
$w_2$	.0717	.0621	.0523	.0434	.0358	.0294	.0230	.0200	.0172

From Cannan, R. K., Kibrick, A., and Palmer, A. H., *Annals N. Y. Acad. Sci.*, **41**, 247 (1941).

It is noted that the effect of increasing salt concentration is to

moderate concentrations; solutions alkaline to this point become more acid. Commonly, the magnitude of the displacement is greatest at reactions far from the isoelectric point. This effect was first clearly revealed by the classic studies of Sørensen<sup>23</sup> upon egg albumin and he recognized that the point of minimum displacement in these curves was identical with, or close to, the isoelectric point of the protein.

TABLE 8. EFFECT OF IONIC STRENGTH (KCl) ON THE DISSOCIATIONS OF  $\beta$ -LACTOGLOBULIN

Ionic strength $I/2$	0.010	0.019	0.035	0.069	0.135	0.270	0.67	2.1
$\bar{Z}$	pH	pH	pH	pH	pH	pH	pH	pH
45							1.90	2.02
44						2.13	2.30	2.42
42				2.10	2.25	2.45	2.70	2.95
40			2.12	2.35	2.53	2.70	2.95	3.16
35		2.36	2.62	2.82	2.96	3.14	3.35	3.52
30	2.65	2.83	3.00	3.19	3.32	3.45	3.67	3.80
25	3.08	3.23	3.36	3.53	3.64	3.74	3.92	4.03
20	3.46	3.60	3.72	3.84	3.93	4.02	4.16	4.25
15	3.86	3.96	4.06	4.15	4.22	4.30	4.40	4.47
10	4.28	4.35	4.42	4.48	4.52	4.57	4.64	4.69
5	4.73	4.75	4.78	4.82	4.85	4.87	4.89	4.92
0	5.18	5.18	5.18	5.18	5.18	5.18	5.18	5.18
-5	5.75	5.72	5.69	5.64	5.63	5.60	5.54	5.51
-10	6.55	6.50	6.45	6.38	6.35	6.30	6.25	6.20
-15	7.55	7.45	7.35	7.28	7.23	7.16	7.10	7.00
-20		8.80	8.65	8.55	8.45	8.40	8.35	8.30
-25			9.72	9.60	9.48	9.40	9.40	9.50
-30			10.20	10.10	10.05	10.02	9.90	10.0
-35			10.60	10.47	10.35	10.30	10.30	
-40			10.83	10.70	10.60	10.55	10.50	
-45			11.05	10.93	10.80	10.75	10.75	
$w$ (Observed)	0.061	0.052	0.046	0.038	0.032	0.027	0.020	0.015
$w$ (Theory)	0.064	0.055	0.047	0.0395	0.0315	0.026	0.020	0.015

From Cannan, R. K., Palmer, A. H., and Kibrick, A. C., *J. Biol. Chem.*, 142, 803 (1942).

These effects in egg albumin were later studied in much greater detail by Sørensen, Linderstrøm-Lang and Lund<sup>1</sup>. Recently, Cannan, Kibrick and Palmer<sup>24</sup> have studied egg albumin over a wide range of ionic strengths in several different salts, and the same authors<sup>25</sup> have made a very thorough study of  $\beta$ -lactoglobulin. Their data for these two proteins, in the presence of KCl at a variety of ionic strengths, are given in Tables 7 and 8. In these tables, the molecular weight of egg albumin has been taken as 45,000, and that of lactoglobulin as 40,000. The mean net charge per mole of protein,  $\bar{Z}$ , is given as a function of pH for different values of  $I/2$ .

The data have been empirically described by Cannan, Kibrick and

Palmer by a simple equation. Choosing the curve at one given ionic strength as a standard, one may find from Table 7 or 8 the difference in pH,  $\Delta\text{pH}_s$ , between the standard and one of the other curves, and plot this as a function of  $\bar{Z}$ . If for egg albumin the curve in 0.067*M* KCl is taken as the standard of reference, it is found that  $\Delta\text{pH}_s$  is a linear function of  $\bar{Z}$  over all regions of the curve.

$$\Delta\text{pH}_s = \alpha_s \bar{Z} + \beta_0 \quad (25)$$

Here  $\alpha$  is the slope of the curve and  $\beta_0$  is a constant factor for each curve, chosen to make all curves intersect at pH 4.9. The  $\beta_0$  term is merely the expression of the fact that the curves at different ionic strength do not intersect at the same isoionic point. The same relation holds for  $\beta$ -lactoglobulin up to  $I/2 = 0.67$ , but in this case  $\beta_0 = 0$ , for all curves intersect at the same isoelectric or isoionic point, pH 5.18.

"The only observable effect of substituting KCl by MCl<sub>2</sub> at constant ionic strength seems to be a parallel displacement of the whole dissociation curve toward a lower pH to an extent increasing with  $I/2$ . In other words,  $\alpha_s$  is unaffected, but  $\beta_0$  is changed in sign. The displacements due to Ca<sup>++</sup> or Mg<sup>++</sup> are almost identical and are slightly greater than those due to Sr<sup>++</sup> or Ba<sup>++</sup>" (reference 24, p. 253).

The theoretical interpretation of these effects is still incomplete but to a first approximation they may be explained in terms of the Debye-Hückel theory. This was first pointed out by Sørensen, Linderstrøm-Lang and Lund<sup>1</sup> and we shall reproduce here the substance of their argument. Consider the dissociation of an acid, A, carrying a charge of  $(Z + 1)$  proton units, to give a proton and the conjugate base B, which carries a charge  $Z$  ( $Z$  may be either positive, zero or negative). Then if  $K$  is the dissociation constant for the reaction we may write, following the discussion in Chapters 3 and 4,

$$\text{pH} - \text{pK} = \log \frac{C_B}{C_A} + \log \frac{\gamma_B}{\gamma_A} \quad (26)$$

where the  $C$ 's denote concentrations and the  $\gamma$ 's activity coefficients of A and B. If now we assume that A and B are spheres and that the net charge of each is distributed uniformly over its surface, we may write, employing equation 183 of Chapter 3:

$$\log \frac{\gamma_B}{\gamma_A} = \frac{\epsilon^2(2Z + 1)}{4.606DRT} \frac{\kappa}{(1 + \kappa a)} \quad (27)$$

Here  $a$  is the "collision diameter" of the protein and the surrounding ions and is slightly greater than the radius,  $b$ , of the spherical protein mole-



Here  $a_A$  is the collision diameter in Å. At very low ionic strengths, the denominator of the term on the right in equation 28 becomes equal to unity, and under these conditions the change in pH with ionic strength, when the ratio of concentrations of acid and conjugate base is held constant, would become

$$\frac{\partial \text{pH}}{\partial \sqrt{I/2}} = 0.5(2Z + 1) \quad (29)$$

Thus if the net charge on the base B is zero or positive, pH increases with increasing ionic strength; if the charge is negative, however, pH decreases. This conclusion is in conformity with the data of Tables 7 and 8, and with all other studies on proteins yet made. Equation 29 further indicates that the change of pH with change in ionic strength should be greater the greater the net charge on the protein. This is again qualitatively in accord with the facts, but the change of pH in solutions far from the isoelectric point is generally much less than would be predicted from a direct application of equation 29.

The actual dissociation of a polyvalent acid or ampholyte may be formulated more precisely, employing a treatment first developed for proteins by Linderström-Lang<sup>26</sup> and further developed by Cannan, Kibrick and Palmer<sup>24</sup>. It may be regarded as a special case of the general treatment given in Chapter 12, equations 37–55 inclusive.

Consider a system of spherical protein molecules, of radius  $b$ , each molecule containing  $m$  carboxyl groups, and  $n$  cationic acid groups, the latter being presumably histidyl, arginyl and lysyl residues. Consider first only the pH region in which all the cationic acid groups carry positive charges. Variations in the net charge on the spherical protein molecule, in this pH range, are thus due only to variations in the number of ionized carboxyl groups. This zone extends from pH 1.5 or 2 nearly to pH 6. In strongly acid solution, the net charge, expressed in proton units, is  $+n$ . If sufficient base has been added to remove  $h$  protons per molecule from the  $m$  carboxyl groups, then the net charge,  $Z$ , is  $n - h$ .

We shall now again make the much oversimplified assumption that the charge  $Z$  is spread uniformly over the surface of the spherical protein molecule. There is some empirical justification for this assumption, because it leads to results which correspond, to a first approximation, to the observed data. Probably this is because there is a set of  $\frac{m!}{(m-h)!h!}$  microscopically different kinds of molecules corresponding to each value of  $h$ , as we have already seen (p. 451). Each individual molecule in the

definite points on or near its surface, and is not at all like a uniformly charged sphere. The average behavior of the entire set, however, is likely to obliterate these individual differences, and may well correspond much more closely to the behavior of a uniformly charged sphere than does that of any individual in the set.

We shall make the further assumption that the intrinsic dissociation tendency of each of the  $m$  carboxyl groups is the same. The probability that a given carboxyl group, at a given pH, has lost a proton, is modified by changes in the electrical potential in its neighborhood. This potential is determined by the net charge on the protein, the dielectric constant, and the ionic strength.

At a given ionic strength in a solvent of dielectric constant  $D$ , the electrical free energy of a uniformly charged sphere of radius  $b$ , with net charge  $(n - h)\epsilon = Z\epsilon$  is, from Chapter 3, equation 130 (neglecting the smaller terms):

$$(F_e)_h = \frac{(n - h)^2 \epsilon^2}{2D} \left( \frac{1}{b} - \frac{\kappa}{1 + \kappa a} \right) \quad (30)$$

The logarithm of the activity coefficient of an ion of this net charge is

$$\ln \gamma_h = \frac{(F_e)_h}{kT} = \frac{(n - h)^2 \epsilon^2}{2DkT} \left( \frac{1}{b} - \frac{\kappa}{1 + \kappa a} \right) = (n - h)^2 w \quad (31)$$

or

$$\gamma_h = e^{(n-h)^2 w} \quad (32)$$

The parameter  $w$ ,

$$w = \frac{\epsilon^2}{2DkT} \left( \frac{1}{b} - \frac{\kappa}{1 + \kappa a} \right) \quad (33)$$

is equal to the potential at the surface of the charged sphere (Chapter 3, equation 120) multiplied by the factor  $\epsilon/(n - h)2kT$ . In a medium of infinite dielectric constant, or for a sphere of infinite radius,  $w$  would become zero. For a sphere of given radius  $b$ ,  $w$  is largest at zero ionic strength, and decreases as the ionic strength increases. For proteins, in which the value of the radius  $b$  is 20 Å or more, the value of  $w$ , even at zero ionic strength, can never be greater than 0.17, and it is much less than this for higher values of  $b$  and  $\kappa$ . The definition of  $\ln \gamma_h$  (equation 31) involves the choice of the standard state as a medium of infinite dielectric constant. This choice is the most convenient in practice for the following discussion.

ions of the class  $A_1$  ( $h = 1$ ); we shall denote it as  $A_{11}$ . The dissociation constant is, denoting concentrations by symbols in brackets:

$$k_{\infty} = (H^+) \frac{[A_{11}]}{[A_0]} \frac{\gamma_1}{\gamma_0} = k'_1 \frac{\gamma_1}{\gamma_0} \quad (34)$$

or, from 31 and 32,

$$k'_1 = k_{\infty} \frac{\gamma_0}{\gamma_1} = k_{\infty} \frac{e^{n^2 w}}{e^{(n-1)^2 w}} = k_{\infty} e^{(2n-1)w} \quad (35)$$

Consider the dissociation of a specified carboxyl group in a particular molecule which has already lost  $h - 1$  carboxyl protons. Denote this molecule as  $A_{(h-1)1}$ , and the conjugate base as  $A_{h1}$ . The dissociation constant is

$$k_{\infty} = (H^+) \frac{[A_{h1}]}{[A_{(h-1)1}]} \frac{\gamma_h}{\gamma_{h-1}} = k'_h \frac{\gamma_h}{\gamma_{h-1}} \quad (36)$$

Here  $k_{\infty}$ —that is, the dissociation constant in a medium of infinite dielectric constant—is the same as in equation 34, since by our hypothesis the intrinsic dissociation tendency of all the groups is the same.

$$k'_h = k_{\infty} \frac{\gamma_{h-1}}{\gamma_h} = k_{\infty} \frac{e^{(n-h+1)^2 w}}{e^{(n-h)^2 w}} = k_{\infty} e^{(2n-2h+1)w} = k'_1 e^{(2-2h)w} \quad (37)$$

The  $\frac{m!}{h!(m-h)!}$  different kinds of molecules of the class with net charge  $Z = n - h$  are all present in equal numbers, since all have the same activity coefficient by equation 32. Hence, denoting by  $(A_h)$  the total concentration of molecules of net charge  $n - h$ , we find:

$$(A_h) = \frac{m!}{h!(m-h)!} (A_{h1}) \quad (38)$$

( $h = 1, 2, 3 \dots n$ ). Thus we may formulate dissociation constants denoted by  $K$ 's, corresponding to the  $K$ 's already defined in equation 5a.

$$K'_1 = (H^+) \frac{(A_1)}{(A_0)} = mk'_1 = mk_{\infty} e^{(2n-1)w} \quad (39)$$

$$K'_h = (H^+) \frac{(A_h)}{(A_{h-1})} = \frac{m-h+1}{h} k'_h = \frac{m-h+1}{h} k_{\infty} e^{(2n-2h+1)w} \quad (40)$$

The total base bound per mole of protein, taking the point of maximum acid binding as the zero of reference, is given by equation 6, replacing the  $L$ 's in that equation by the corresponding  $K$ 's defined in 5a.

$$\frac{K'_1}{(1+K'_1)} + \frac{2K'_1 K'_2}{(1+K'_1)^2} + \dots + \frac{mK'_1 K'_2 \dots K'_m}{(1+K'_1)^m}$$

and the  $K$ 's are defined by 39 and 40. Thus  $\bar{h}$  is a function of  $k_{\infty}$ , ( $H^+$ ),  $m$  and  $w$ . The carboxyl group titration curves ( $\bar{h}$  - pH curves) defined by 41 are qualitatively similar to the curves already defined by equation 8. The interaction between the groups, determined by the factor  $w$ , spreads the curve over a wider pH zone. Like the curves defined by 8, however, the curves defined by 41 are symmetrical about their mid-point, if  $\bar{h}$  is plotted against pH. Near the mid-point, the curve is approximately a straight line; and if  $m$  is large, the molal buffer value in this region was found by Linderstrøm-Lang<sup>21</sup> to be

$$\frac{d\bar{h}}{dpH} = -\frac{d\bar{Z}}{dpH} = \frac{2.303m}{2(wm + 2)} \quad (42)$$

Thus if the slope of the titration curve at the mid-point, and the number,  $m$ , of carboxyl groups is known, the value of  $w$  can be determined. Values of  $w$  obtained for egg albumin from the titration curves at various ionic strengths are listed in Table 7 in the row headed  $w_{\text{exp.}}$ . Here  $m = 51$ , assuming a molecular weight of 45,000. These values have been compared with values calculated from equation 33, taking  $D$  as the dielectric constant of water, and  $b$  as 27.5 Å, the value obtained from the assumed molecular weight and the partial specific volume of egg albumin. Two values for  $a$  in equation 33 were considered: (1)  $a = b = 27.5$ ; (2)  $a = b + 2 = 29.5$  Å. The value 2 Å was taken as the mean radius of the  $K^+$  and  $Cl^-$  ions. The values of  $w$  corresponding to these two choices of  $a$  are listed in Table 7 as  $w_1$  and  $w_2$  respectively. Neither is identical with  $w_{\text{exp.}}$ , but the ratio  $w_{\text{exp.}}/w_2$  is consistently  $0.81 \pm 0.01$  for all values of  $I/2$  below 0.7. For  $\beta$ -lactoglobulin (Table 8) the values of  $w_{\text{exp.}}$  and  $w_2$  are almost identical. From the analysis of the titration curve of lactoglobulin the value of  $m$  was taken to be 58 carboxyl groups per mole, and the radius  $b$  was calculated from the diffusion constant\* as 29.1 Å. Equation 40 may be written in logarithmic form as

$$pK'_h = pk_{\infty} - \log \frac{m - h + 1}{h} - \frac{(2n - 2h + 1)w}{2.303} \quad (43)$$

In actual calculations Cānnan, Kibrick and Palmer set  $w = w_{\text{exp.}} = 0.8w_2$  for egg albumin, and  $w_{\text{exp.}} = w_2$  for  $\beta$ -lactoglobulin.

It is interesting to note the influence of the fixed positive charge,  $n$ , on the  $pK'_h$  values. If  $n'$  of these positive charges are removed—as, for instance, by deamination of the  $\epsilon$ -ammonium groups of lysine, without otherwise altering the protein—then the entire carboxyl portion of the titration curve should be displaced along the pH axis by  $2n'w/2.303$  pH

Since the  $\bar{h}$  - pH curves, for the carboxyl groups, are symmetrical about their mid-points, the pH of the mid-point of the carboxyl titration is equal to  $\text{pK}'_{\frac{m+1}{2}}$ . It follows from 43 that

$$\text{pH}_{\text{mid.}} = \text{pk}_{\infty} - \frac{(2n - m)w}{2.303} \quad (44)$$

Employing this equation and the data of Table 7, consistent values of  $\text{pk}_{\infty}$  (carboxyl) =  $4.30 \pm 0.04$  were found for egg albumin<sup>24</sup> for all ionic strengths from 0.0085 to 2.38. Equally consistent results were obtained for  $\beta$ -lactoglobulin<sup>25</sup>, but  $k_{\infty}$  was found to be 4.6. No structural basis is known for this difference of 0.3 pH units between the values for the two proteins.

The computation of the titration curves from equation 41 is an extremely long and tedious process. Fortunately a simple relation has been found which permits the computation of the actual curves, corresponding to any given value of  $w$ , to a very good approximation. When  $w = 0$ , the curve obtained is identical with that of  $m$  equivalents of a univalent acid for which  $\text{pK} = \text{pk}_{\infty}$  (see equations 1 and 8). When a set of curves corresponding with various positive values of  $w$  are compared with a curve defined by 8, it is found that the former are displaced from the latter on the pH-axis by an amount approximately equal to  $0.868wZ = 0.868w(n - \bar{h})$ <sup>27a</sup>. That this should be true in the central linear portion of the curve is implicit in equation 42. In practice the same relation is found to apply with relatively high precision over the whole course of a series of curves corresponding to a range of values of  $w$  much wider than those expected to apply to proteins. This leads to the following approximate dissociation equation for a set of  $m$  ionizing groups which all have the same intrinsic dissociation constant.

$$\text{pH} = \text{pk}_{\infty} - 0.868w(n - \alpha m) - \log [(1 - \alpha)/\alpha] \quad (44a)$$

Here  $\alpha$  represents the fraction of all the groups of this class which have lost protons, that is,  $\alpha = \bar{h}/m$ . At the mid-point of the curve ( $\text{pH}_{\text{mid.}}$ )  $\alpha = 0.5$ . Consequently

$$\text{pH} = \text{pH}_{\text{mid.}} - 0.868wm(\alpha - 0.5) - \log [(1 - \alpha)/\alpha] \quad (44b)$$

These equations are found to reproduce the curves more precisely defined by equation 41, very closely indeed. The curves given by equations 44 never deviate from those calculated from 41 by more than 0.1 equivalent per mole of polyvalent acid.<sup>27a</sup> It is doubtful whether the titration curve of any protein has ever been reproduced as accurately as this.

or the  $\epsilon$ -ammonium groups of lysine<sup>28</sup>. Practically the same values for  $w$  were found in egg albumin from the analysis of the slope of the titration curve near pH 11, with 23 ammonium groups involved, as from the analysis of the carboxyl titration near pH 4;  $pK_{\infty}$  for the ammonium groups was found to be 10.07, and for the imidazolium groups 6.7<sup>24</sup>. In  $\beta$ -lactoglobulin, the corresponding value was 6.8, employing the value of  $w$  used in describing the carboxyl titration. The mid-points and slopes of the amino titration, on the other hand, corresponded to much smaller values of  $w$ .

The theory outlined here is open to criticism on at least three grounds: (1) The distribution of charge on the protein ions is certainly not uniform—the charges are discrete, and distributed according to some pattern. This point has already been discussed. (2) Proteins are in general not spherical,\* although many may be nearly so. (3) The carboxyl groups are of at least two kinds, aspartyl and glutamyl residues; hence there are at least two different values of  $pK_{\infty}$  (carboxyl), and not one as the theory assumes. The two  $pK_{\infty}$  values, however, should not differ by more than a few tenths of a pH unit. It is difficult to estimate how seriously this factor should affect the calculated curves.

In view of the extreme simplification involved, it is remarkable that the theory fits the facts as well as it does. It is necessarily a first approximation, due to be superseded in time by a more adequate picture of the actual situation. It is apparent, however, that even the simple theory serves to interpret successfully a large body of complex experimental facts.

### The Titration Curves of Certain Important Proteins

Having considered certain general principles underlying the interpretation of protein titration curves, we may now turn to the detailed analysis of the titration curves of certain individual proteins. The first published titration curve of a protein was apparently given by D'Agostino and Quagliariello<sup>29</sup>, who studied serum albumin in 1912. Their data were obtained only over a restricted pH range, but they attempted to estimate the dissociation constants of the groups in the protein from the slope of the titration curve. Their results, which were confirmed and extended by later investigators, marked the development of a new phase in the study of the physical chemistry of proteins. Since that time a vast body of data has accumulated, describing the titration curves of proteins and their combining capacity with acids and bases. In 1925 Cohn<sup>2</sup> gave a comprehensive discussion of the data then available, and eight years later Pauli

<sup>28</sup> *ibid.*, p. 100. The value of  $w$  is the same near pH 5.5 in which the dissociation of carboxyl and imida-

and Valkó<sup>30</sup> gave a very thorough survey of the available data. In the following discussion we shall restrict ourselves to a limited group of proteins. Some of these are chosen for discussion because the proteins themselves are very well characterized chemical individuals—for instance, hemoglobin, cytochrome C, egg albumin,  $\beta$ -lactoglobulin, and serum albumin. Others, like gelatin and wool keratin, are not well defined chemical individuals but are discussed because of the general significance of some of the experiments which have been carried out upon them.

**Hemoglobin.** Among well known proteins, probably the hemoglobin of the horse has been studied in more different ways than any other. First of all, we may consider the titration curve of horse CO-hemoglobin, as determined by Cohn, Green and Blanchard<sup>16</sup>, between pH 1 and 13. This is represented graphically in Fig. 1. The curve in Fig. 1 is constructed as though hemoglobin gave up  $261 \times 10^{-5}$  mole of hydrogen ions per gram of protein between pH 1.5 and 13. The maximum combining capacity with base is given in Table 9.

TABLE 9. ACID AND BASE COMBINING CAPACITY OF HEMOGLOBINS

Protein	Acid combining capacity moles $\times 10^5$ per gm protein	Base combining capacity moles $\times 10^5$ per gm protein	Total combining capacity moles $\times 10^5$ per gm protein
Ox oxyhemoglobin (1)	145		
Horse carboxyhemoglobin (2)		105	
(3)	159	124.5	283.5
(4)	148	113	261

(1) Lewis, P. S., *Biochem. J.*, **21**, 46 (1927).

(2) Stone, G. C. H., and Failey, C. F., *J. Phys. Chem.*, **37**, 935 (1933).

(3) Pauli, W. and Schwarzscher, W. in Pauli and Valkó, "*Kolloidchemie der Eiweisskörper*," Theodor Steinkopff, Leipzig, 1933.

(4) Cohn, E. J., Green, A. A., and Blanchard, M. H., *J. Am. Chem. Soc.*, **59**, 509 (1937).

**Maximum Acid and Base Combining Capacity.** Several studies of the combination of hemoglobin with acid and base have been reported. One of these is the titration with the quinhydrone electrode by Lewis<sup>31</sup> of the carboxyhemoglobin of the ox with sulfuric acid in the presence of ammonium sulfate. His results yield an acid combining capacity of  $145 \times 10^{-5}$  equivalent per gram of hemoglobin. Pauli and Schwarzscher<sup>32</sup> report that hemoglobin combines  $159 \times 10^{-5}$  mole hydrochloric acid per gram and  $124.5 \times 10^{-5}$  mole lithium hydroxide per gram. The results of Stone and Failey<sup>33</sup>, based on solubility measurements on thallos chloride in alkaline hemoglobin solutions, yield a base combining capacity of slightly over

<sup>30</sup> Pauli, W. P., and Valkó, E., "*Kolloidchemie der Eiweisskörper*," Dresden and Leipzig, Theodor Steinkopff (1933), pp. 31-88 inclusive.

$100 \times 10^{-5}$  mole per gram calculated on the basis of the activity coefficient of thallos chloride and the value of 14 for  $pK_w$ .

Cohn, Green and Blanchard<sup>16</sup> obtain a maximum acid combining capacity of  $148 \times 10^{-5}$  mole per gram on horse CO-hemoglobin, in close agreement with the results obtained by Lewis on ox hemoglobin and only slightly below the values of Pauli and Schwarzacher. The agreement between the various authors may therefore be regarded as satisfactory.

All these values are appreciably higher than the estimates for the total histidine, arginine and lysine content as given by Vickery<sup>34</sup>, the latter amounting to  $125.3 \times 10^{-5}$  mole per gram of hemoglobin. The analytical data are all based on the amounts actually isolated and some of them may therefore be low. The agreement between the analytical and titration data here is by no means satisfactory. For globin, after removal of the prosthetic heme groups, Cohn, Salter and Ferry<sup>35</sup> found a lower acid binding capacity of  $138 \times 10^{-5}$  mole per gram, a value somewhat closer to the analytical figures for hemoglobin.

The value for the maximum base binding by hemoglobin is a more uncertain quantity for reasons indicated earlier in this chapter. Cohn, Green and Blanchard<sup>16</sup> estimated this quantity as of the order of magnitude of  $113 \times 10^{-5}$  mole per gram of hemoglobin, but the scatter of the individual points alkaline to pH 12 is great, and it is uncertain whether all of the guanidino groups of arginine were titrated at this pH. The relation between the analytical data for hemoglobin and acid and base combining capacity is not as satisfactory as for some of the other proteins discussed in this chapter. However, the relation between the titration constant and the structure of the dissociating groups has been carried out in detail by several quite different methods, which give results in excellent accord.

**The Titration Constants of Hemoglobin.** The whole titration curve given in Fig. 1 is readily divisible into at least three clearly distinct regions, one extending from saturation with acid roughly to pH 5, the second from pH 5 to 9 and the third from pH 9 to the most alkaline reactions studied. If this curve is analyzed on the basis of the smallest possible number of titration constants, the results obtained may be described as follows. Taking the molecular weight of hemoglobin as 67,000, each free reactive group combines with  $1.5 \times 10^{-5}$  mole of acid or base per gram. In Table 10 the titration constants used in describing the data of the titration curve are listed along with the number of groups assigned to each titration constant. The probable nature of these groups is also indicated.

Comparison of these assignments with the available analytical data brings out several points of great interest. The number of carboxyl groups



87 per molecule as deduced from the titration curve, is more than three times as great as the number accounted for by the analysis of Chibnall and Bailey, as given by Vickery<sup>34</sup>. They report 45 residues of aspartic acid and 29 of glutamic acid per mole. From these must be subtracted 48 residues of amide nitrogen per mole which presumably bind an equal number of dicarboxylic acid residues in the protein molecule. This leaves only 26 free carboxyl groups per molecule of hemoglobin on the basis of analytical data. It is obvious that there is a very large discrepancy here. It will be seen later that the same is true of other proteins such as gelatin and egg albumin.\* On the other hand, the number of residues with titration constants at  $pK$  5.7 and 7.5 is found to be 33, and this is identical with the value deduced from the histidine determinations of Vickery and Leavenworth. Comparison of the most alkaline  $pK$  values at 10.8 and 11.6 with the analytical data is more difficult. The best available analyses indicate the presence of 37 lysine residues, 14 arginine residues and 12 tyrosine residues per molecule of hemoglobin—a total of 63. Fifty-four residues

TABLE 10. THE TITRATION CONSTANTS OF HORSE CO-HEMOGLOBIN IN THE ABSENCE OF ADDED SALT

$pK_1'$	$pK_2'$	$pK_3'$	$pK_4'$	$pK_5'$	$pK_6'$	$pK_7'$
3.7	4.0	4.8	5.7	7.5	10.8	11.6
Number of groups of each constant dissociating						
88	4	13	20	40	14	
Carboxyl groups		Histidine		Lysine and arginine or tyrosine		

per mole are accounted for in this range on the basis of the titration curve. It is practically certain that some at least of the guanidine residues in arginine are not titrated even at the most alkaline reactions reached. Cohn, Green and Blanchard conclude that "The titration curve does not reveal a sufficient number of dissociable groups in the range alkaline to pH 9 to account for arginine, lysine and tyrosine, but enough to account for lysine and either tyrosine or arginine."

The titration constants given in Table 10 are for hemoglobin in the absence of added salt. In one molal sodium chloride the titration constants are markedly different. The curves are rotated around the isoelectric point in the manner already described for egg albumin and lactoglobulin. At points acid to the isoelectric point, the curve is displaced to more alkaline reactions by the addition of salt, while the reverse is true alkaline to the isoelectric point. The titration curve in the presence of 1*M* sodium chloride (Fig. 1) is constructed using the values of 5.2, 6.1, 7.1, 10.4 and 11.2 for  $pK_3'$ ,  $pK_4'$ ,  $pK_5'$ ,  $pK_6'$  and  $pK_7'$ , respectively;  $pK_3'$  and

$pK'_4$  are thus 0.4 greater, and  $pK'_5$ ,  $pK'_6$  and  $pK'_7$  are 0.4 less than the values in Table 10.

The form of the titration curve at reactions acid to pH 4 as shown in Fig. 1 must be regarded as an artifact. In this pH range the addition of acid is breaking down the combination between heme and globin, thereby altering the titration constants of some of the groups studied, and perhaps simultaneously releasing some new titratable groups. Therefore, it is found that the back titration with alkali of a solution of hemoglobin already saturated with acid gives a quite different curve in this range, the form of the curve being much flatter than that indicated in Fig. 1. Therefore the titration constants ascribed to the carboxyl groups in Table 10 are not to be taken as indicative of the actual dissociation constants of these groups in the hemoglobin molecule. This limitation, however, does

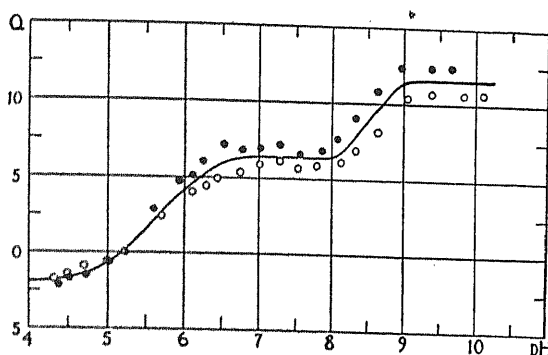


Figure 2. Apparent heat of dissociation of oxyhemoglobin as a function of pH  
 ● = values for the 25-37.7° interval; ○ = values for the 6.5-25° interval.  
 From Wyman, J., Jr., *J. Biol. Chem.*, **127**, 1 (1939).

not invalidate the calculation in Table 10 of the number of carboxyl groups involved in the titration curve of hemoglobin nor of the total limiting value of the maximum acid combining capacity.

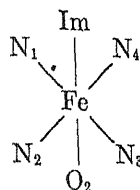
**Heat of Ionization of the Acid and Basic Groups in Hemoglobin.** Wyman<sup>36</sup> has determined the heat of ionization of the acidic groups in the horse oxyhemoglobin molecule with  $pK$  values between 4 and 10 by titrating the hemoglobin at temperatures of 6.5°, 25° and 37.5°. A quantity which may be called the apparent heat of dissociation,  $Q'$ , may be calculated from such curves by analogy with the case of a weak monobasic acid, employing the equation:

$$Q' = -RT^2 \left( \frac{\partial \log K}{\partial T} \right)_{pH}$$

In this equation,  $R$  denotes the gas constant,  $T$  the absolute temperature and  $B$  the total amount of base present in the system per mole of hemoglobin. When  $Q'$  is plotted against pH, the curve shown in Fig. 2 was obtained by Wyman. This curve shows three horizontal zones, one acid to pH 5, one extending approximately from pH 6 to pH 8 and one extending from pH 9 to more alkaline reactions. Within any one of these flat regions, the ordinate  $Q'$  on the curve indicates the apparent heat of dissociation of the groups ionizing in that pH range. The groups of Class I, those ionizing below pH 4.5, have an apparent heat of dissociation of  $-2000$  or  $-3000$  calories. This value is within the range characteristic of the heat of ionization of carboxyl groups. Groups in Class II, ionizing between pH 6 and 8, have an average apparent heat of dissociation of about  $+6200$  calories. This value is entirely consistent with that found as characteristic of the imidazole nucleus in histidine and its peptides. Groups in Class III, ionizing at pH alkaline to 9.5, have an average apparent heat of dissociation of  $+11,500$  calories. This figure is within the range characteristic of the  $\epsilon$ -ammonium groups of lysine, but is much higher than the value of about  $+6000$  characteristic of the phenolic group of tyrosine.

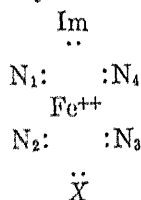
There are two transition zones between these three horizontal regions of the curve, one zone centering about pH 5.5, the other about pH 8.5. The analysis of the transition from the ionization of a group of one type to that of another type involves a more complex theoretical treatment; fully given in Wyman's original communication. This beautiful thermal analysis of the titration curve of hemoglobin carried through by Wyman leads unequivocally to conclusions regarding the nature of the ionizing groups which are completely in accord with the interpretation of Fig. 1 as given by Cohn, Green and Blanchard.

**The Heme-Linked Acid Groups of Hemoglobin.** The distinctive properties of hemoglobin are due to the combination of globin with the attached prosthetic group, the iron porphyrin complex, heme. Each of the four heme groups in the hemoglobin molecule contributes two carboxyl groups to the acid groups of hemoglobin. Moreover, at least one amino acid residue of globin is directly linked to the iron atom of each heme; and at least one other such residue is very close to the iron, although in oxy or CO-hemoglobin, not directly linked to it. Recent advances in the chemistry of hemoglobin have provided a fairly definite picture of the structure binding heme to globin<sup>37</sup>. The configuration of groups around the iron atom is similar to that in the ferrocyanide ion,  $\text{Fe}(\text{CN})_6^{\equiv}$ , in which the



Here the four pyrrole nitrogen atoms ( $\text{N}_1\text{N}_2\text{N}_3\text{N}_4$ ) of the porphyrin nucleus lie, surrounding the iron atom in the plane of the paper; they form a square with the iron at its center. The other two groups are situated directly above and below the iron atom. The group binding the iron to the globin is denoted as Im, since much evidence, discussed further below, indicates that it is an imidazole residue of histidine.

The formula of CO-hemoglobin is similar to that of oxyhemoglobin, the bonds holding the iron atom to the six surrounding groups being covalent in both cases. The latter point was demonstrated by Pauling and Coryell<sup>38</sup>, who showed that these molecules are diamagnetic<sup>39</sup>. They made the surprising discovery, however, that ferrohemoglobin (reduced hemoglobin)<sup>40</sup> is markedly paramagnetic, the magnetic moment found corresponding to four unpaired electrons per heme. This indicates that the bonds holding the iron atom to the surrounding groups are essentially ionic: they may be represented by the formula



The dots represent unshared electrons. The group X is a sixth group, whose presence was inferred by Conant<sup>37</sup> on various grounds. In oxyhemoglobin, Conant assumed that X remains close to the iron atom but is prevented by the attached oxygen from bonding with it directly. The titration data, discussed below, suggest that X is a second histidine imidazole group. The introduction of the oxygen atom in the sixth coordination position around the iron atom, and the radical change in bond type associated with this process, give rise to important changes in the acidity of the groups Im and X.

It has long been known that under physiological conditions oxyhemo-

<sup>38</sup> Pauling, L., and Coryell, C. D., *Proc. Nat. Acad. Sci.*, 22, 159 (1936).

<sup>39</sup> Cf. "Electromagnetic Criterion for Bond Types," see Pauling, L., "The Nature of the Chemical Bond," Ithaca, N. Y., 1939, pp. 112-123 inclusive.

globin is a stronger acid than ferrohemoglobin, and conversely pH has a profound influence on the affinity of hemoglobin for oxygen<sup>41</sup>. Recently German and Wyman<sup>42</sup> reinvestigated this problem using the glass electrode for their titrations. Their results, like those of earlier workers, revealed that between the pH values of 6.1 and 9.0 more base was bound at a given pH by oxyhemoglobin than by hemoglobin. They also observed an effect not previously recognized, namely, that between pH 4.5 and 6.1 (the "acid loop" of the curve) oxyhemoglobin binds less base than hemoglobin. Above pH 9 and below pH 4.5, on the other hand, no observable difference was found between the two forms. German and Wyman con-

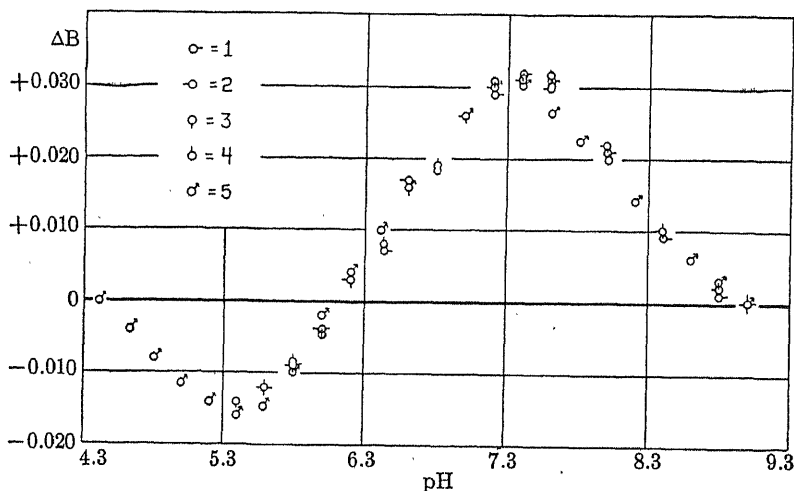


Figure 3. Difference in base bound (milli equivalents per gram) between oxygenated and reduced hemoglobin, as a function of pH, for Experiments 1 to 5 given in Tables I to V of German and Wyman, Ref. 42.

From German, B., and Wyman, J., Jr., *J. Biol. Chem.*, **117**, 533 (1937).

cluded, therefore, that "in the alkaline loop the effect of oxygenation is to increase the acid dissociation constants of one or more oxylabile groups, while in the acid loop, on the other hand, its effect is to diminish them. This implies that at reactions between 4.5 and 6.1 the familiar Bohr effect of the alkaline range (according to which oxygen affinity increases with pH) should be reversed. Actually, Ferry and Green<sup>43</sup> found that this effect is reversed below a pH of about 6.5" (42, p. 540).

The difference in the amount of base bound at a given pH between oxy- and ferrohemoglobin is plotted as a function of pH in Fig. 3.

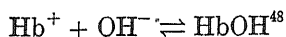
<sup>41</sup> The most comprehensive survey of this field up to 1928 is given in the treatise of T. J. Henderson, "Blood

The effect of ionic strength on the difference between the two forms of hemoglobin at a given pH was found to be extremely small.

Recently, Wyman<sup>44</sup> has given a more searching analysis of this phenomenon by studying the difference in the titration curves of oxy- and ferrihemoglobin at three different temperatures, 7°, 25°, and 38°. From the changes in the titration curves with change in temperature he deduced the variation in the heat of oxygenation of hemoglobin with pH, throughout the entire pH range from 3 to 11. "The results can be accounted for on the basis of the dissociation of hydrogen ion which accompanies oxygenation if we assume a heat of dissociation of 6500 calories per equivalent. This is the heat of dissociation characteristic of the imidazole group of histidine, and indicates that it is groups of this kind which dissociate as a result of oxygenation of the hemoglobin molecule." . . . "The heat is the same for each stage of the oxygenation process, as is the shift in the amount of base bound at constant pH. A further analysis based on the shift in the amount of base bound at 25° indicates that each molecule of hemoglobin contains four identical heme-histidine complexes, in which two histidine units interact with each heme, the energy of interaction amounting to 680 calories for one and to 1380 calories for the other" (reference 44, p. 598). At 25° the pK' values of the two histidines are 5.25 and 7.81 in ferrihemoglobin; the corresponding values in oxyhemoglobin are 5.75 and 6.80. The acidity of the former group is decreased, that of the latter increased, by oxygenation. The group with the higher pK value shows the greater interaction energy with iron and oxygen; presumably it corresponds to Im in the formula given above; the other group would then correspond to X.

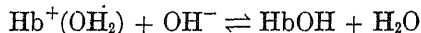
Similar heme-linked acid groups may be discerned in ferrihemoglobin (methemoglobin). Taylor and Hastings<sup>45</sup> showed that "upon the oxidation of hemoglobin to methemoglobin, for each electron transferred there appears a univalent cation corresponding to an apparent acidic dissociation constant (K) of  $2.2 \times 10^{-7}$  (pK = 6.65)" (45, p. 661). The existence of another such acid group in ferrihemoglobin, with pK = 5.25, was inferred by Coryell and Pauling<sup>46</sup> from titration curves, oxidation reduction potentials, and magnetic susceptibility measurements; they concluded that these two pK values (5.25 and 6.65) corresponded to the two heme-linked imidazole groups. In addition, however, ferrihemoglobin shows a third heme-linked acid group not found in oxy- or ferrihemoglobin. This group is clearly revealed by a pronounced change in absorption spectrum with pH, occurring chiefly in the pH region between 7 and 9. The equilibrium between the two forms was studied spectrophotometrically by Haurowitz,

and by Austin and Drabkin<sup>47</sup>. Their results could be described in terms of the reaction:



Austin and Drabkin evaluated the pK for this reaction (*i.e.*, the pH at which  $[\text{Hb}^+] = [\text{HbOH}]$ ) as 8.12 at ionic strength 0.10, for dog hemoglobin, while the earlier work of Haurowitz shows the existence of a group in horse ferrihemoglobin, with pK 8.2. Coryell, Stitt and Pauling<sup>49</sup> studied the same group in bovine hemoglobin by measuring magnetic susceptibility.  $\text{Hb}^+$ , like Hb, is strongly paramagnetic, its magnetic moment indicating the presence of five unpaired electrons per heme (ionic bonds between iron and the surrounding atoms).  $\text{HbOH}$  is much less strongly paramagnetic (probably three unpaired electrons). Determinations of magnetic susceptibility as a function of pH gave  $\text{pK}' = 8.15$  at ionic strength 0.2, and 8.56 at ionic strength 1.3. Recently Wyman and Ingalls<sup>50</sup> have determined this  $\text{pK}'$  as 8.05 for horse ferrihemoglobin, by means of a differential electrometric titration of ferri- and oxyhemoglobin.

This acid group is either the Fe atom itself or a water molecule directly bonded to it<sup>49, 46</sup>. In the former case, Fe acts as an acid in that it binds the base, hydroxyl ion, although it is not a proton donor. In the latter case, the attached water molecule would act as a proton donor



Magnetic studies on ferrihemoglobin hydroxide,  $\text{HbOH}$ , ferrihemoglobin fluoride,  $\text{Hb}^+\text{F}^-$ , and other ferrihemoglobin compounds with cyanide, hydrosulfide, azide, and other compounds have been carried out by Coryell, Stitt and Pauling<sup>49</sup> and Coryell and Stitt<sup>51</sup>; and Coryell and Pauling<sup>46</sup> have given an ingenious structural interpretation of the pK values of the heme-linked acid groups in *ferro*-, *ferri*- and oxyhemoglobin<sup>52</sup>. Michaelis and Granick<sup>53</sup> have shown that the magnetic moment of catalase is very similar to that of  $\text{HbOH}$ .

**Cytochrome C.** Another heme-protein, which in contrast to hemoglobin contains only one heme group per molecule, is cytochrome C. Its properties are in many respects very different from those of hemoglobin. It does not combine reversibly with oxygen at any pH; the *ferro* form combines with CO, but only at strongly acid or alkaline reactions. The iron, how-

<sup>47</sup> Austin, J. H., and Drabkin, D. L., *J. Biol. Chem.*, **112**, 67 (1935); Haurowitz, F., *Z. Physiol. Chem.*, **138**, 68 (1924).

<sup>48</sup> Here  $\text{Hb}^+$  denotes the ferrihemoglobin molecule containing one iron atom (the whole molecule being  $(\text{Hb}^+)_4$ ); ferric iron has five unpaired electrons (electron per heme).

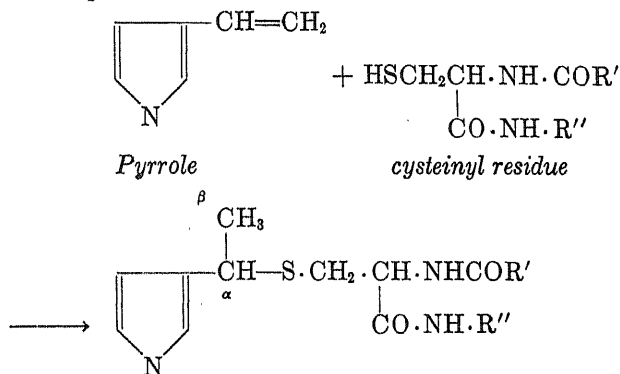
<sup>49</sup> Coryell, C. D., Stitt, F., and Pauling, L., *J. Biol. Chem.*, **139**, 633 (1937).

<sup>50</sup> Wyman, J., Jr., and Ingalls, C. V., *J. Biol. Chem.*, **139**, 633 (1937).

<sup>51</sup> Coryell, C. D., and Stitt, F., *J. Biol. Chem.*, **139**, 633 (1937).

<sup>52</sup> Coryell, C. D., and Stitt, F., *J. Biol. Chem.*, **139**, 633 (1937).

ever, is readily oxidized from the *ferro* to the *ferric* state; in the living cell the oxidation of cytochrome a, b, and c is generally brought about by another heme-protein, cytochrome oxidase<sup>54</sup>. Cytochrome C is one of the smallest of all well-characterized proteins. After the final electrophoretic purification a uniform product is obtained, containing 0.43% iron and 1.48% sulfur. The molecular weight, assuming one Fe and six S atoms per molecule, is 13,000<sup>55</sup>. The heme is bound to the protein by bonds between heme iron and amino acid nitrogen, as in hemoglobin. Also apparently there are linkages between the two vinyl radicals, attached to pyrrole rings in the protoporphyrin residue of the heme, and cysteine residues of the protein



There should be two such links in the cytochrome molecule<sup>56</sup>. Because of these linkages, it is possible for both of the bonds between the protein and the heme iron to be broken without disrupting the heme. This appears to occur in strongly acid solution ( $\text{pH} < 1$ ), as judged by spectrophotometric and magnetometric measurements<sup>57</sup>. In neutral solution, however, the heme iron appears to be completely coordinated with the four surrounding pyrrole nitrogen atoms and two amino acid nitrogens. The latter, as in hemoglobin, appear to belong to imidazole residues of histidine.

The amino acid composition of cytochrome C is radically different from that of hemoglobin. The analyses of Theorell and Åkesson<sup>55</sup> indicate the presence of two arginyl, three histidyl and approximately 22 lysyl residues per mole. The lysine content is the highest yet reported for any protein. The amino nitrogen content of the unhydrolyzed protein is 31-32 atoms amino-N per mole, much higher than the lysine content estimated by isolation procedures.

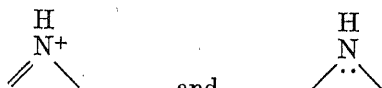
<sup>54</sup> Warburg, O., *Ergeb. Enzymforschung*, 7, 210 (1937); Keilin, D., and Hartree, E. F., *Proc. Roy. Soc. (London)*, B125, 171 (1938); Stern, K. G., *Ann. Rev. Biochem.*, 9, 1 (1940); Theorell, H., *ibid.*, 9, 663 (1940).



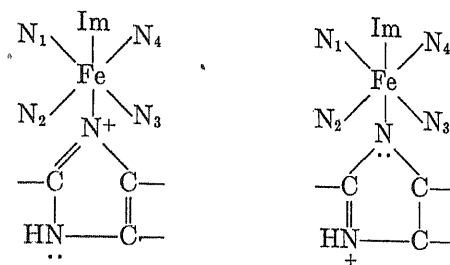
A similar discrepancy has been found in egg albumin, in which isolation procedures yield about 36 moles of lysine per  $10^5$  grams of protein, while the results of the formaldehyde titration and amino nitrogen determinations agree in giving a much higher figure, between 48 and 57 moles of amino nitrogen per  $10^5$  grams (exclusive of 2 or 3 groups per  $10^5$ , presumed to arise from  $\alpha$ -amino nitrogen)<sup>58</sup>. A discrepancy of the same order of magnitude has recently been reported also in  $\beta$ -lactoglobulin<sup>25</sup>. It must be remembered, however, that amino nitrogen determinations, as ordinarily carried out, often give falsely high results for certain amino acids, notably glycine and cystine. Amino nitrogen determinations on proteins may be subject to similar errors, but the discrepancy found in cytochrome C is so large that it probably cannot be entirely explained on this basis. The same applies to insulin, for which Chibnall<sup>77</sup> has found eighteen free amino groups per mole, in excess of the lysine  $\epsilon$ -amino nitrogen.

The number of titratable carboxyl groups has not yet been very exactly fixed. Theorell and Åkesson<sup>59</sup>, however, determined the titration curve at both 0 and 20°, and found  $Q'$  (equation 45) to be about  $-1000$  cal/mole at pH values acid to 5.3, and to rise steeply and steadily at more alkaline reactions to a value of about  $+10,000$  beyond pH 9. About eighteen carboxyl groups per mole are titrated between pH 5.2 and the most acid solutions measured. The number of these groups is far smaller than the number of basic amino acid residues, consequently the isoelectric point is very alkaline, being at pH 10.09 for *ferri*-, and 10.17 for *ferrocytochrome* at 20°.

Cytochrome C, in striking contrast to hemoglobin, contains very little histidine. Study of the titration curves at 0 and 20° indicated that only one histidyl residue, out of the three determined analytically, is titrated in the pH range 5.5-8.5. The other two histidines may be accounted for by considering the difference in the titration curves of *ferri*- and *ferrocytochrome*. In the pH range 8-10.6 two equivalents more are titrated in the former than in the latter. Theorell<sup>57b</sup> has concluded that the data are best explained by assuming that the imino groups of the imidazole rings are titrated in this region, one with  $pK'$  9.35, one with  $pK'$  9.85. These are of course quite different from the  $pK'$  values commonly found for the imidazole nucleus, but the underlying process is also quite different. The free imidazolium ion ( $pK$  near 7) has a formula which may be derived from resonance between structures such as



There are two  $\text{—NH}$  groups, from either of which dissociation may take place. On the other hand, the imidazole rings bound to the iron atom in cytochrome have such resonating structures as



(Here again  $N_1$ ,  $N_2$ ,  $N_3$ ,  $N_4$  denote the four pyrrole nitrogens of the porphyrin, and Im denotes the second imidazole nucleus attached to the iron.) In this case there is only one acidic hydrogen attached to an imidazole nitrogen atom, and the conditions for dissociation are entirely different from those in the free imidazolium-group. The  $pK'$  values of 9.35 and 9.85 for the two imidazole groups attached to the iron appear very plausible in the light of the finding of Russell and Pauling<sup>60</sup> that the covalent imidazole ferrihemoglobin complex has a  $pK'$  value of 9.5. They attribute this  $pK'$  value to the  $\text{=NH}$  group in the imidazole.

The study of the dissociation constants of other heme proteins, such as myoglobin, catalase, peroxidase and cytochrome oxidase offers a rich field for future investigation. Also the study of simple hemochromogens, formed by the union between heme and simple nitrogenous compounds, should yield illuminating results, which may serve to clarify our understanding of the behavior of the heme proteins.

**Zein.** The titration curve of zein is of particular interest because this substance is the only one of the prolamines whose titration curve has hitherto been studied in some detail. It is not feasible to carry out the titration in water, since zein at its isoelectric point and even in strongly acid solution is insoluble in this solvent. Cohn, Berggren and Hendry<sup>61</sup> reported an estimate of the maximum base combining capacity of zein in water but were unable to observe any combination with acids in this solvent. The latter observation, however, was due simply to the insolubility of the zein, since in alcohol-water mixtures its acid-combining capacity was found to be measurable, although far lower than that of almost any other known protein. The figures in Table 11 show the acid- and base-combining capacity of the trivalent amino acids in zein. The titration curve of zein in alcohol-water mixtures was reported by Cohn, Edsall

calculation of the bound and free acid were carried out in a manner exactly corresponding to that employed for aqueous solutions. The results as obtained between pH values of 3 and 9 in 80% ethanol are graphically reported in Fig. 4.

TABLE 11. ACID AND BASE COMBINING CAPACITY OF TRIVALENT AMINO ACIDS OF ZEIN

	moles per gram $\times 10^5$		moles per gram $\times 10^5$
Glutamic acid (a).....	212.7	Histidine (e).....	5.3
Aspartic acid (b).....	13.5	Arginine (f).....	10.5
Oxyglutamic acid (c).....	15.3	Lysine (g).....	0.0
Dicarboxylic acids (a + b + c) ..	241.5	Dibasic acids, B <sup>+</sup> (e + f + g)...	15.8
Ammonia (d).....	214.1		
Free dicarboxylic acids, A <sup>-</sup> = (a + b + c - d) ..	27.4	Excess acid groups, A <sup>-</sup> - B <sup>+</sup> ..	12.1
		Total free groups, A <sup>-</sup> + B <sup>+</sup> ..	43.7
Tyrosine.....	32.6		

Vickery, H. B., *Compt. rend. trav. lab. Carlsberg*, **22**, 519 (1938), has obtained 9.1 moles of arginine, and 5.0 moles of histidine, per  $10^5$  gm zein, by very careful isolation procedures. These values are slightly lower than earlier values of Kossel and Kutscher, tabulated above. The data in this table are taken from Vickery's paper.

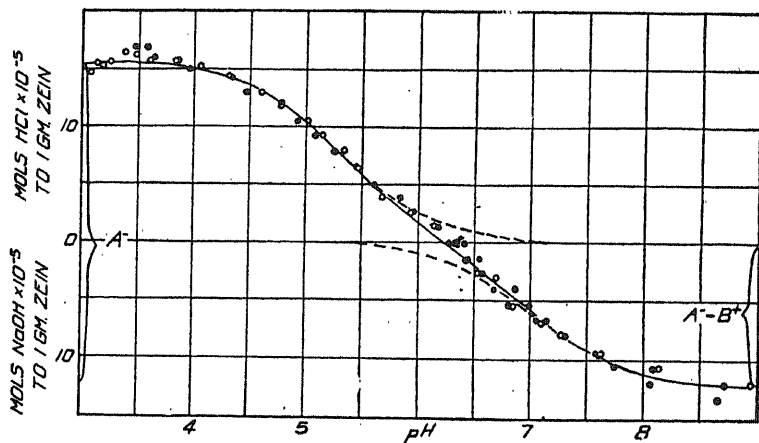


Figure 4. Titration curve of zein in 80 per cent alcohol.  
From Cohn, E. J., Edsall, J. T., and Blanchard, M. H., *J. Biol. Chem.*, **105**, 319 (1934).

Two different zein preparations were studied which gave slightly different acid combining capacities. Preparation O combined with  $17.8 \times 10^{-5}$  mole of acid per gram of zein, while the acid combining capacity of Preparation B was  $21.3 \times 10^{-5}$  mole per gram of zein. These values are only slightly in excess of the analytical determinations on the dibasic acids of zein as

extends from the most acid solutions studied to approximately pH 8. This portion of the titration curve can be accurately described in terms of two constants,  $pK_1 = 5.4$  and  $pK_2 = 6.9$ . Beyond pH 8 there is a zone of small buffer value. Following this point there is a zone in which considerable combination with alkali occurs. Measurements upon the different preparations did not agree satisfactorily in the alkaline range and Cohn, Edsall and Blanchard therefore did not report their results on this portion of the titration curve.

An extensive investigation of the titration curve of zein and of iodozein was reported by Neuberger<sup>63</sup>. He showed that the effect of temperature

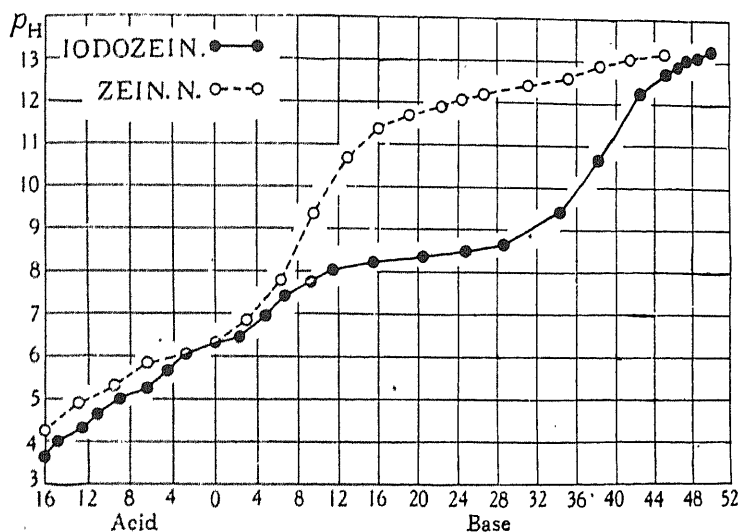


Figure 5. Acid and base binding of zein and iodozein.  
From Neuberger, A., *Biochem. J.*, **28**, 1990 (1934).

was in the direction to be expected from the nature of the groups ionizing in the various pH ranges. Solutions more acid than pH 5.5 showed very little change in pH between 25 and 45°. Between pH 5.5 and 7, the pH values were lower by 0.30–0.35 units, while in alkaline pH ranges the difference was much larger, about 1 pH unit. These differences are quite comparable to those found by Wyman for hemoglobin, allowing for the fact that the dissociation of carboxyl groups is shifted in alcohol-water mixtures to considerably larger pH values than those characteristic of aqueous solutions.

Neuberger's most interesting results, however, were obtained by a com-

titration curves is shown in Fig. 5. Acid to pH 7 there is little difference between the two curves, but between pH 8 and 13 they deviate profoundly, approaching one another again in strongly alkaline solutions. The steepest portion of the titration curve in Fig. 5 is shifted by about 3.5 pH units to the acid side as compared to the curve of zein. This is very nearly equal to the difference between the pK values of the hydroxyl group in tyrosine and diiodotyrosine. Neuberger concluded both from the titration curve and from the iodine analysis of iodozein that iodine was introduced only into the benzene ring of tyrosine. While Neuberger's evidence for zein appears to justify this conclusion, the results obtained on the treatment of other proteins with iodine have led to far more complex results<sup>4, 64</sup>. There is evidence that histidine can react with iodine to form a diiodo derivative, and Philpot<sup>3</sup> has concluded that still other groups in certain proteins of a nature as yet unknown can also react with iodine. Zein contains a very large amount of tyrosine and very little histidine; the reverse is true in the case of hemoglobin. In the light of these differences in composition, the radical difference in the effect of iodine on the titration curves of these two proteins might perhaps have been expected. While no simple interpretation of the study of hemoglobin can be offered at the present, and no general interpretation of the titration curves of iodinated proteins can therefore be made in accordance with any simple rule, the behavior of zein is readily interpretable in terms of formation of diiodotyrosine residues.

**The Titration Curve of Gelatin: Influence of Formaldehyde and of Alcohol.** The titration curve of gelatin has been studied by a number of authors; notably, and with particular care, by Hitchcock<sup>65</sup>. He determined the maximum combining capacity for acid as  $96 \times 10^{-5}$  mole per gram of a standard gelatin preparation; the isoelectric point was 4.85. The curve is steep from the extreme acid region to about pH 6; between pH 6 and 9, it is relatively flat. The total number of groups titratable between pH 1 and 8 is  $132 \times 10^{-5}$  mole per gram; in the region between pH 9 and 12, about  $54 \times 10^{-5}$  more is titrated, but it is probable that maximum base binding has not been reached at pH 12. Different preparations of gelatin, studied by various investigators, show significant differences in acid and base binding capacity (reference 2, p. 383). This is not surprising in view of the drastic procedures commonly employed in preparing gelatin from collagen, and the variety of sources from which commercial gelatin is obtained.

Further consideration of the titration curve of gelatin is desirable, however, because of the variety of techniques employed by various investigators to elucidate the nature of the groups ionizing in various pH ranges.

The effect of formaldehyde is notable (see Chapter 5). Birch and Harris<sup>66</sup> titrated gelatin in water and in 1% formaldehyde, and showed that the most acid part of the titration curve was practically unaffected by the formaldehyde, while alkaline to pH 5 the apparent pK values of the dissociating groups were shifted to more acid regions. These results were in accord with the results obtained by these authors on amino acids and peptides. Recently the effect of formaldehyde on gelatin has been studied

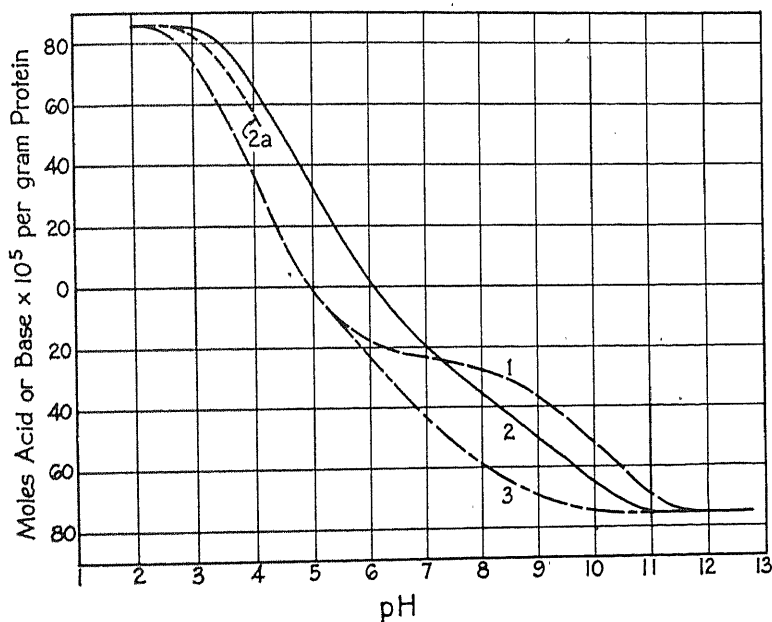


Figure 6. The titration curve of gelatin (1) in water, (2) in 80% ethanol-water mixtures containing 0.02 *M* KCl, and (3) in 1% formaldehyde solution. Acetate and glycinate buffers were employed as titrating reagents over the larger part of the curve. The method of calculating bound acid or base is described in the original communication. Curve 2a is derived from measurements in 80% ethanol-water, without added KCl. Based on data of Lichtenstein, I., *Biochem. Z.*, **303**, 20 (1940).

in more detail by Lichtenstein<sup>67</sup>, in H. H. Weber's laboratory. Her results are shown in Fig. 6. The curves for water and for 1% HCHO are identical up to pH 5.5, and diverge widely at more alkaline reactions; it was inferred that the break at this point represents the end of the region where carboxyl groups are being titrated, and the beginning of the titration of imidazole groups. Between pH 1.5 and 5.5,  $96 \times 10^{-5}$  mole per gram of gelatin was titrated—a figure somewhat lower than that found by

lytical figures (Chapter 15) for free carboxyl groups in the dicarboxylic acids of gelatin account for only  $42 \times 10^{-5}$  mole per gram of gelatin. The great discrepancy suggests that the isolation of dicarboxylic acids after hydrolysis, even in the most careful studies hitherto made, is still very far from complete. The same conclusion has already been reached in the discussion of hemoglobin<sup>68</sup>. Fig. 6 also shows (solid points) the titration curve of gelatin in 80% alcohol (+0.02*M* KCl). The acid portions of the curve are displaced toward more alkaline reaction in this medium, as compared with water; the alkaline portions of the curve are displaced toward more acid reactions. This is in harmony with the effects of alcohol in the titration curves of amino acids (Chapter 4).

**Effects of Deamination.** In 1923 Hitchcock<sup>69</sup> carefully deaminated gelatin at low temperatures, and found that the product of deamination had lost  $40 \times 10^{-5}$  equivalent of nitrogen per gram, as compared with the original gelatin. This figure corresponds closely to the lysine content as determined by analysis. After deamination the isoelectric point was shifted to pH 4.0, and the maximum acid binding capacity had diminished from 89 to  $44 \times 10^{-5}$  mole acid per gram. The difference of  $45 \times 10^{-5}$  corresponds reasonably well with the loss of nitrogen as determined by analysis.

Lichtenstein's recent studies also include deaminized gelatin (Fig. 7). Hitchcock's findings were confirmed and other significant results obtained. The number of groups titrated between pH 1 and 8 is identical, before and after deamination, within the limits of experimental error. The groups titrated between pH 8.5 and 12, on the other hand, are almost completely removed by deamination. This is the region in which the  $\epsilon$ -amino groups of lysine would be expected to titrate. The effects of alcohol (Curve 2) on the titration curve of deaminized gelatin are entirely in accord with the assumption that the free lysine ammonium groups have been removed, leaving the other groups unaltered. (The anomalous curve 2a in Fig. 7 was obtained in 80% alcohol, without added salt; curve 2 in the presence of 0.02*N* KCl. The peculiar form of curve 2a is probably due to liquid junction potentials varying markedly with pH; addition of low concentrations of KCl eliminates the disturbance).

Deamination appears to have little effect on the pH range in which the carboxyl groups of gelatin give up protons to base. The number of carboxyl groups ionized at any pH acid to 8 (taking the region of maximum acid binding as the zero of reference) is nearly the same in Figs. 6 and 7.

<sup>68</sup> Birch and Harris (66) have shown a marked effect of formaldehyde in low concentrations on  $pK_2$  of histidine; it may therefore be inferred that a similar effect on the  $pK$  of imidazole groups in proteins should be found, and the data shown in Figure 6 are best interpreted in this light.

This indicates that the  $\epsilon$ -ammonium groups of lysine in gelatin are in general so far from the carboxyl groups that the removal of positive charge from the lysines does not affect the COOH dissociation significantly<sup>70</sup>. This is not what would be expected from the application of equation 43 or 44, from which a considerable increase of the carboxyl pK values after deamination would be inferred. Equation 43, however, is derived for a system of protein molecules which are monodisperse and spherical. Gela-

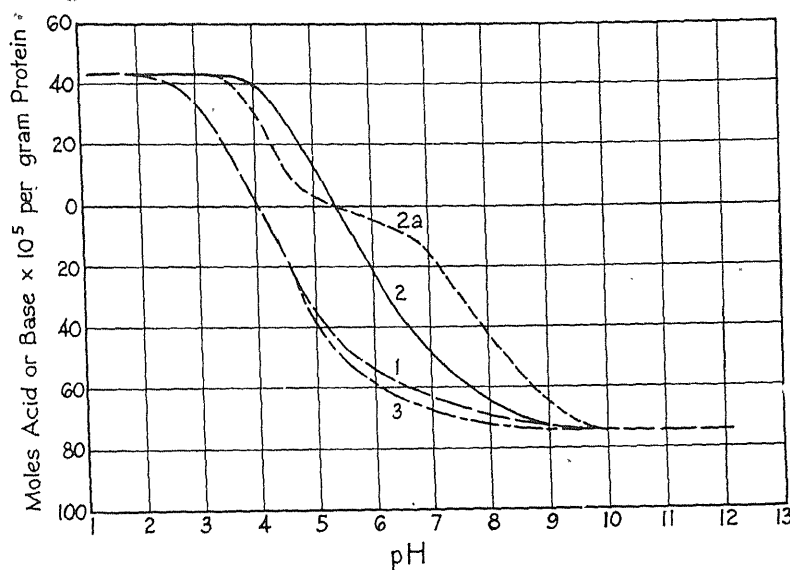


Figure 7. The titration curve of deaminized gelatin (1) in water, (2) in 80% ethanol-water mixtures containing 0.02 M KCl, and (3) in 1% formaldehyde solution. The procedures employed were the same as for the data in Fig. 6. The anomalous curve, 2a, was obtained in 80% ethanol-water, without added KCl. The anomalous form of the curve disappeared on adding salts to the system (curve 2). Based on data of Lichtenstein, I., *Biochem. Z.*, **303**, 22 (1940).

tin preparations are highly polydisperse, and the molecules are certainly not spherical. Therefore the attempt to apply equation 43 to gelatin is probably quite unjustified.

#### Wool Protein: Effect of Different Acid Anions on the Titration Curve.

A detailed study of the reaction of wool protein (which is largely keratin) with hydrochloric acid and potassium hydroxide was carried out by Stein-

<sup>70</sup> Lichtenstein points out that her figure for the maximum acid binding of gelatin ( $56 \times 10^{-5}$  mole acid/mole gelatin) is significantly lower than the theoretical  $50 \times 10^{-5}$  mole acid/mole gelatin.



hardt and Harris<sup>71</sup>. The wool fibers are found to take up neither acid nor base from the surrounding solution when the pH is 6.3–6.5 at 0°. The maximum acid combining capacity is 82 moles per  $10^5$  gm of wool and is independent of the ionic strength. The maximum base binding capacity is greater than 78 moles per  $10^5$  gm. As the ionic strength increases, the carboxyl portion of the titration curve is displaced toward more alkaline reaction and the ammonium group portion of the curve near pH 11 is displaced toward more acid reactions. Qualitatively therefore the effect of ionic strength on the titration curve of wool is like that already described for egg albumin and lactoglobulin (p. 468). The magnitude of the shift, however, is apparently somewhat greater than in the case of the latter proteins. Steinhardt and Harris have described their data in terms of the hypothesis that anions as well as hydrogen ions are bound by the proteins in the acid portion of the titration curve. It would appear, however, that the data might also be described by an electrostatic theory similar to that employed by Linderstrøm-Lang<sup>21</sup> and Cannan<sup>24, 25</sup>. It would be obviously wrong, however, to use a spherical protein model to describe the reactions of the wool fiber with acid and base. The simplest model to employ in this instance might be a uniformly charged cylinder of very great length. The electrical potential in the neighborhood of such a cylinder in a medium of given dielectric constant could perhaps be employed in calculations analogous to those already given for the spherical model.

The effect of temperature on the titration curve of wool protein<sup>72</sup> is qualitatively similar to that on all other proteins investigated. In the carboxyl portion of the titration curve the apparent heat of ionization is approximately +2500 calories per mole. In the ammonium portion of the curve it is of the order of magnitude of 13,000 calories per mole. A correction factor should be introduced for the heat of transfer of the hydrogen ion from the aqueous solution surrounding the fiber into the fiber itself, since this process must precede the actual reaction of the hydrogen ion with the basic groups in the wool. This correction factor has been estimated as of the order of magnitude of 2500 calories per mole<sup>72</sup>. When this figure is subtracted from the heats of ionization given above, values almost identical with those found for other proteins are obtained.

A very interesting study has been made of the relative affinities of the anions of a large number of different strong acids for wool protein<sup>73</sup>. The pH values obtained for a given number of equivalents of acid added per gram of wool vary very greatly indeed from one acid to another. The form of the titration curves obtained is well shown in Fig. 8, taken from a comprehensive review of this work which has lately been presented by

Steinhardt<sup>74</sup>. The pH value at which 40 millimoles of acid are bound per 10<sup>2</sup> grams of protein varies from about 2.3 in the case of hydrochloric acid to 4.2 in the case of flavianic acid. The general form of all the curves is

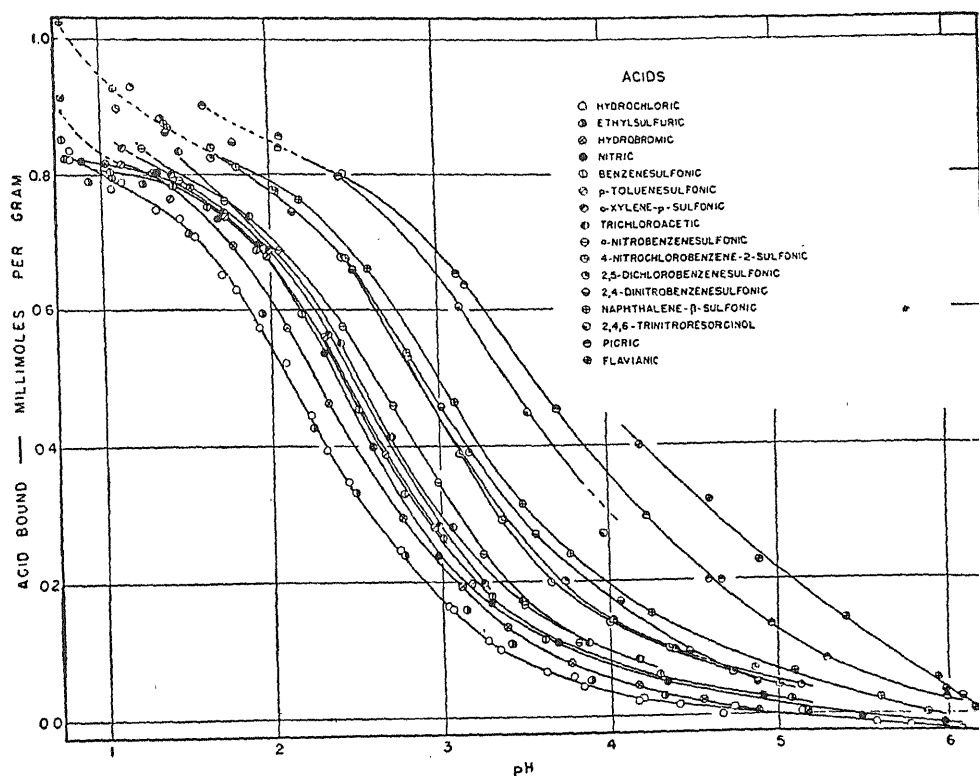


Figure 8. Combination of wool protein with sixteen different strong acids as a function of pH at 0°C.

The measurements with flavianic acid are represented in milliequivalents instead of millimoles. Points representing the acids o-xylene-p-sulfonic, and o-nitrobenzenesulfonic) would fall at lower pH values than are required for congruence with the curves for the other acids, and have been omitted to avoid confusing the figure. Data for the two dibasic acids, flavianic and trinitroresorcinol, are only partially represented.

Tables of the data included in this figure are published in *J. Res. Nat. Bur. Standards*, **26**, 293 (1941).

From Steinhardt, J., *Ann. New York Acad. Sci.*, **41**, 287 (1941).

very similar. They differ from each other primarily by parallel displacements along the pH axis in the same direction.

and the anions of the various acids. The greater the affinity of the protein for the acid anion, the more the midpoint of the curve is displaced towards larger pH values. The different acid anions studied show such a high degree of individuality that it is difficult to lay down any general rules correlating relative affinity with the structure of the acid in question. Steinhardt has, however, in his discussion pointed out certain significant factors involved. Similar effects were found by Steinhardt when egg albumin was titrated with a number of the same acids shown in Fig. 8. The relative displacements of the curves, however, were found to be far smaller for the soluble protein egg albumin than for the insoluble wool protein. The difference between the curves obtained with hydrochloric and flavianic acids, for instance, was only about 0.6 pH units when 40 moles of acid had been bound per  $10^5$  grams of egg albumin, as compared with 2 pH units in the case of wool protein. It appears probable, therefore, that the insoluble fibrous proteins undergo this type of reaction to a much larger extent than the soluble proteins. Reactions of this sort have of course long been employed in technology, for example, in such processes as the tanning of leather, in which the combination of protein with the anion of tannic acid is undoubtedly an important factor.

**Egg Albumin.** Crystalline egg albumin has been known as a pure protein over a very long period of time. A relatively crude preparation of egg albumin was studied by Bugarszky and Liebermann<sup>11</sup> in 1898, and some of their data have already been presented in Table 2. A series of measurements by later investigators were plotted by Cohn<sup>2</sup> in 1925 and showed generally good agreement over the major portion of the curve. The value estimated at this time for the maximum acid combining capacity was 80 moles per  $10^5$  grams of protein; but more recent investigations lead to a slight increase of this value. Kekwick and Cannan<sup>75</sup> estimated the maximum acid combining capacity as  $89 \pm 2$  moles per  $10^5$  grams of protein. The published curve of Lichtenstein<sup>67</sup> indicates a value near 85 moles per  $10^5$  grams of protein, but this value is probably less exact. Recent data from this laboratory<sup>76</sup> on 5 times recrystallized egg albumin (44.28 grams protein per liter) give the figure of  $87.2 \pm 2$  for measurements in the pH range 1.8–1.4. Some of these measurements were carried out with no added salt; others in the presence of 0.15 molal sodium chloride. The most recent report of Cannan, Kibrick and Palmer<sup>24</sup> leads to the slightly higher value of  $91 \pm 2$ . The general agreement of these results from three different laboratories may be regarded as satisfactory.

Cannan, Kibrick and Palmer<sup>24</sup> have expressed their results in terms of an assumed molecular weight of 45,000 for the egg albumin molecule. The

studying the effect of formaldehyde and of temperature. The net charge of the molecule near pH 8.5 in water, as determined from the base bound per mole, represents the negative charge of all the carboxyl groups minus the cations present at this pH. In water the latter will be the sum of the amino and guanidino groups. In the presence of formaldehyde the guanidino groups will be the only cations at this pH. Thus it is possible from the formaldehyde titration to calculate a figure for the carboxyl groups minus the guanidino groups. The number of histidine imidazole groups may be determined from the temperature coefficient of the titration curve and is found to be five per mole ( $\pm 1$ ).

A well defined break in the titration curve of egg albumin is found at pH 8-8.5. In aqueous solution this corresponds to a net charge, in proton units per mole, of  $-15 \pm 0.5$ , while in the presence of excess of formaldehyde<sup>58</sup> the value is  $-37 \pm 0.5$ . The difference between the net charge in water and in formaldehyde at pH 9 is 22. This is a measure of the amino groups which react with formaldehyde. The titration data thus lead to the following estimate of the numbers of the various ionizable groups in egg albumin:

(a) Total cations.....	41
(b) Carboxyl minus guanidine.....	37
(c) Amino.....	22
(d) Imidazole.....	5
(e) $a - c - d =$ guanidine.....	14
(f) $b + e =$ carboxyl.....	51

Egg albumin has been reported to contain roughly one mole of phosphoric acid. Such a group would be expected to contribute two equivalents to the curve. One of these would be included in the carboxyl region near pH 2 and the other in the imidazole region near pH 7. The estimate of carboxyl and imidazole groups given above would include any phosphoric acid present.

The analytical data on the dibasic and dicarboxylic acids in egg albumin expressed on the basis of the molecular weight 45,000 are as follows (Chapter 15, Table 4):

Residue	Ionizable group	Equivalents per mole
(a) Lysine.....	$\epsilon$ -Ammonium	15
(b) Histidine.....	Imidazolium	4
(c) Arginine.....	Guanidinium	14
(d) Aspartic acid.....	Carboxyl	28

The values for arginine and histidine derived from analysis and from the titration curve are in very satisfactory agreement. The lysine content, on the other hand, is insufficient to account for all the groups which react with formaldehyde. The reaction of nitrous acid with egg albumin essentially confirms the formaldehyde titration, since it gives a value of 23 to 24 amino groups per mole<sup>58</sup>.

Earlier analytical work gave much lower yields of aspartic and glutamic acids than those listed here; and a large discrepancy between carboxyl groups, as determined by analysis and by titration respectively, was pointed out by Lichtenstein<sup>67</sup>. The beautiful analysis of Chibnall<sup>77</sup>, on which the figures given above are based, has practically removed this discrepancy, and indicates that the titration data for carboxyl groups give a valuable indication of the result to be expected from careful analytical work.

**$\beta$ -Lactoglobulin.** The crystalline globulin isolated by Palmer<sup>78</sup> from the albumin fraction of cow's milk has recently been designated as  $\beta$ -lactoglobulin; and its titration curve has been reported in detail by Cannan, Palmer and Kibrick<sup>25</sup>. An extensive tabulation of their data has already been given in Table 8. The isoionic point was found to be  $5.18 \pm 0.01$ , in dilute sodium chloride or potassium chloride, and the addition of potassium chloride up to  $2m$  did not significantly change this pH. This value is in excellent agreement with the electrophoretic isoelectric point of  $5.19$ <sup>79</sup>. The titration curve was found to be reversible from pH 2-10, the back titration curve within this pH range being identical with the original curve at the same ionic strength. In the neighborhood of pH 11 there are indications of irreversible reactions which become very rapid at greater alkalinity. At pH 12  $\beta$ -lactoglobulin is converted in less than five minutes to a product insoluble in dilute salt solutions in the isoelectric region. The titration curve of this altered product differs significantly from that of the unaltered protein. The ultracentrifugal studies of Pedersen<sup>79</sup> indicate that the protein suffers an irreversible breakdown above pH 11. The effect of temperature and of formaldehyde on the titration curve were studied in exactly the same manner as in egg albumin and the analytical data inferred directly from the titration curve are as follows, expressed on the basis of a molecular weight of 40,000 for the protein:

(a) Total cations.....	46
(b) COOH — guanidine .....	52-53
(c) COOH — guanidine — amino .....	18-19
(d) Imidazole.....	6

The analytical data are as follows (Chibnall<sup>77</sup>, assuming a molecular weight of 40,000).

Residue	Ionizable group	Equivalents per mole
(a) Histidine.....	Imidazolium	4
(b) Lysine.....	$\epsilon$ -Ammonium	27
(c) Arginine.....	Guanidinium	7
(d) Aspartic acid.....	Carboxyl	30
(e) Glutamic acid.....	Carboxyl	59
(f) Ammonia.....		30
(g) Free carboxyls (d + e - f).....		59

Here again the values for histidine and arginine from titration agree very well with those determined by isolation procedures. The lysine value is again low and the same comments apply here as in the cases of egg albumin and cytochrome C. The dicarboxylic acids determined from Chibnall's<sup>77</sup> analyses are in most excellent agreement with the results to be expected from the titration data.

**Insulin.** An important study of the titration curve of insulin was reported in 1936 by Harington and Neuberger<sup>80</sup>. Over a large pH range near its isoelectric point, insulin is so insoluble that no titration curve can be obtained, but in more acid solutions Harington and Neuberger obtained an acid binding capacity of approximately 100 moles per  $10^5$  grams of protein. This value is in very satisfactory agreement with recent determinations from our own laboratory, which are shown in Fig. 9. This total is somewhat higher than the value of 85.6 obtained for the dibasic amino acids by isolation procedures (Chapter 15, Table 4). Insulin, like hemoglobin, is rich in histidine, of which 51.6 moles per  $10^5$  grams of protein have been isolated. Insulin binds 63-66 moles per  $10^5$  grams per mole of protein between the isoelectric point and pH 8.5, according to Harington and Neuberger. The figure obtained in this laboratory is only a few per cent lower. Tyrosine analyses indicate the presence of 68 moles tyrosine per  $10^5$  grams insulin, or approximately 24 moles of tyrosine per mole of insulin, assuming a molecular weight of 35,000. Harington and Neuberger found that the iodine content of fully iodinated insulin corresponds to 24 tyrosine groups per mole, on the assumption that the iodine has entered exclusively the 3:5 positions of tyrosine residues. The titration curve of iodinated insulin over the pH range 6.6-11.5 showed a displacement toward more acid reaction as compared with that of normal insulin. The character of the displacement was almost exactly similar to that found in

This is similar to that of all other proteins investigated in showing a very slight displacement in the region where the carboxyl groups are titrated and a much larger displacement in the region in which imidazole and amino groups are titrated. The calculated values for the heat of ionization of the amino groups as deduced from this curve are somewhat lower than those found in most other proteins. The exact form of the curves shown is to

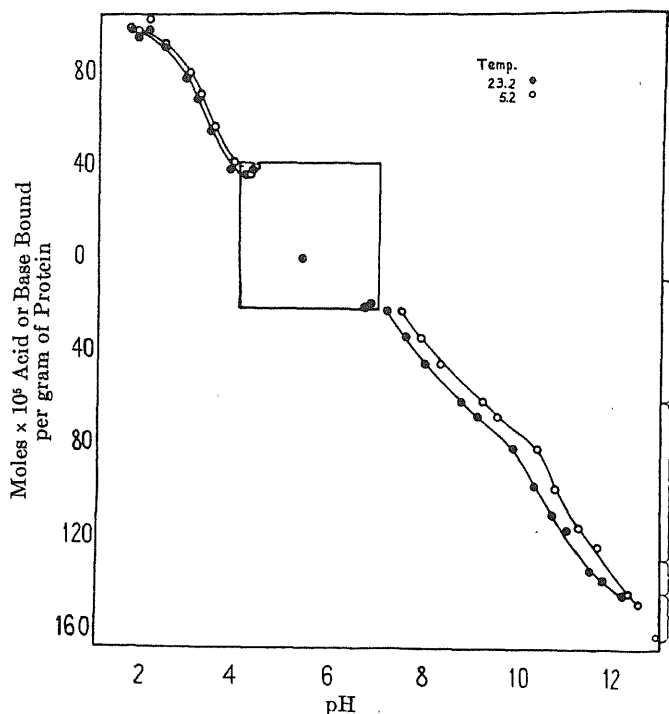


Figure 9. The titration curve of insulin at two different temperatures. Insulin becomes insoluble over the pH range enclosed in the square within the figure and therefore scarcely any points have been obtained within this range. Measurements of Cohn, E. J., Ferry, J. D., and Blanchard, M. H.

be regarded as tentative and may require future revision; but their general character is clear.

**Horse Serum Albumin.** The serum albumin of the horse was the first protein for which a titration curve was determined<sup>29</sup> and repeated studies have been made upon it. The curves of several of the earlier workers were in very good agreement, as was shown in 1925 by a composite plot of their data (reference 2, p. 380). Unfortunately, the exact significance of these

shows that all earlier preparations consisted of a mixture of proteins. Fig. 10 shows the titration curve of a crystalline carbohydrate-free preparation of horse serum albumin which was homogeneous with respect to molecular weight and electrophoresis studies. The values plotted in the

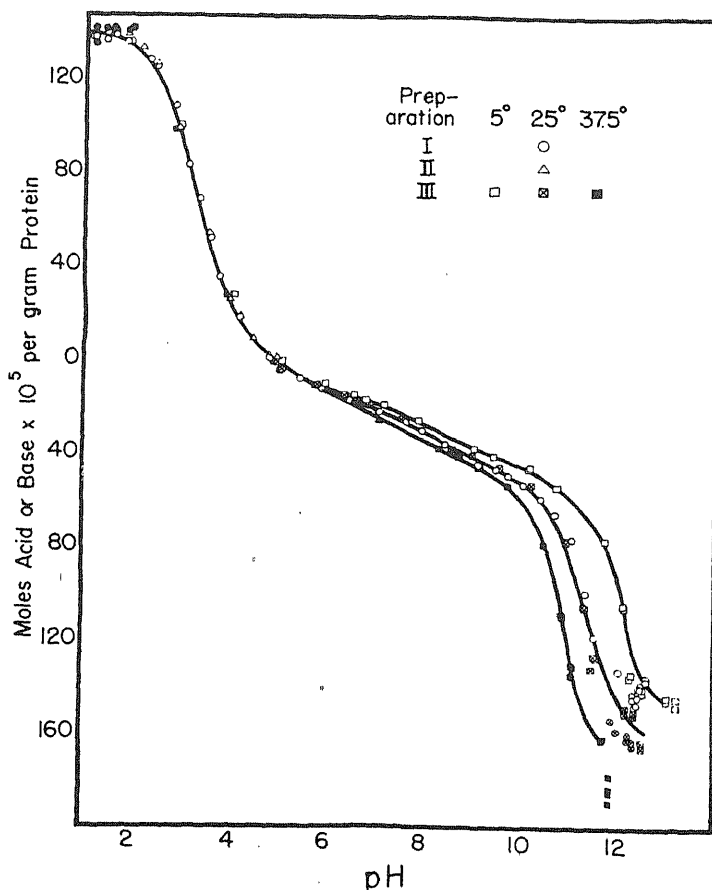


Figure 10. The titration curve of crystalline horse serum albumin at three different temperatures. Measurements of Cohn, E. J., Strong, L. E., and Blanchard, M. H.

curve were obtained on three different preparations, two prepared by Dr. T. L. McMeekin and one, following McMeekin's method, by Dr. A. C. Batchelder. The results on the three preparations, the last of which was studied at three different temperatures, are in very satisfactory agreement.



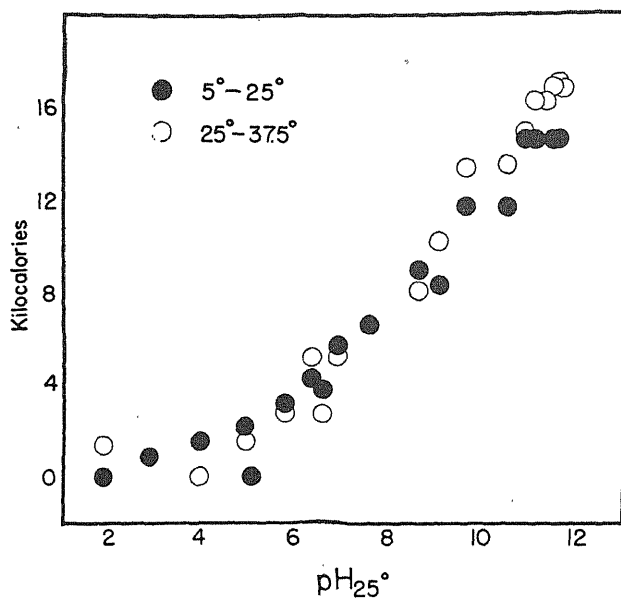
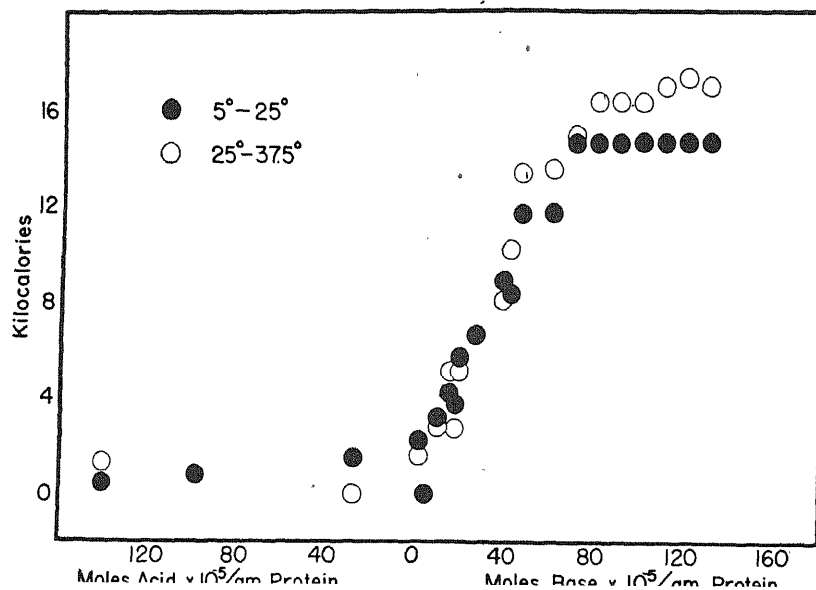


Figure 11. The heat of ionization of horse serum albumin plotted as a function of pH. Measurements of Cohn, E. J., Strong, L. E., and Blanchard, M. H.



reported earlier by several workers on unfractionated serum albumin (reference 2, p. 380). The base bound at pH 12 on the preparation of McMeekin is close to 160 per  $10^5$  grams protein. In the measurements at  $37.5^\circ$  near pH 12, large fluctuations in base binding were observed from one experiment to another, depending particularly on the time during which the protein had been exposed to the alkaline solution. Clearly at this temperature and pH irreversible reactions are proceeding with considerable rapidity and the three low points in the curve for  $37.5^\circ$  are to be given no significance.

The apparent heat of ionization  $Q'$  (equation 45) has been plotted against the pH value at  $25^\circ$  for horse serum albumin in Fig. 11. The values of  $Q'$  for the carboxyl groups at pH  $< 5$  are between 0 and 2000 calories per mol, as in other proteins. Between pH 6–9 there is a steep, fairly steady rise with only a very slight indication of a plateau corresponding to the imidazole groups. At very alkaline reactions between 11–12  $Q'$  rises above 14,000 calories per mol. This is a very high value even for the heat of ionization of ammonium groups; it is possible that secondary irreversible reactions are affecting the calculated  $Q'$  value at this pH. This is particularly likely to be true for the  $Q'$  values calculated from the measurements at  $37.5^\circ$ .

Fig. 12 shows  $Q'$  data plotted against the moles of acid or base bound by the protein instead of the pH. This method of representing the data has an advantage over the other in that the width of the region in the curve corresponding to groups of any particular type is proportional to the number of groups involved. It appears, (see also<sup>36</sup>) that there is considerable overlapping of groups in the range in which the imidazole groups of histidine are being titrated, and that the measured value of  $Q'$  in the transition region between the carboxyl and amino portions of the titration curve is determined by several different types of groups ionizing practically simultaneously.

## Chapter 21

### Rotary Brownian Movement. The Shape of Protein Molecules As Determined from Viscosity and Double Refraction of Flow\*

BY JOHN T. EDSALL

The study of sedimentation and diffusion gives information concerning the shape of proteins (Chapter 18, equations 27 and 28) as well as their size. The dissymmetry of protein molecules is commonly calculated as if the protein were an ellipsoid of revolution, characterized by the relative lengths of the semi-axis of revolution ( $a$ ) and the equatorial semi-axis ( $b$ ). The values of  $a/b$  so obtained are subject to some uncertainty, owing to the unknown degree of hydration of the protein; and also owing to uncertainty as to whether the simplified ellipsoidal model used is adequate to represent a substance as complex as a protein.

The importance of the problem of protein shape makes it desirable to utilize all available methods of approach to this problem. There are several methods which are quite independent of sedimentation and diffusion measurements. In this chapter we shall discuss the determination of viscosity and double refraction of flow in protein solutions, and their interpretation; and in Chapter 22 that of dielectric dispersion measurements. All these measurements involve the rotation and partial orientation of protein molecules in an external field of force. In dielectric dispersion measurements the orienting force is produced by an alternating electric field; in viscosity and double refraction of flow, rotation is produced by shearing forces arising from velocity gradients in the streaming liquid. In any case, however, the orientation achieved is only partial, since it is opposed by the disorienting action of the Brownian movement of the molecules. The Brownian movement here to be considered is not that of translation discussed in Chapter 18 in connection with translational diffusion. It is an analogous phenomenon involving rotation of the molecules about their

orienting forces<sup>1</sup>. In the presence of such orienting forces, a steady state is gradually achieved, a state intermediate between the two limiting conditions of complete orientation and of complete disorder. The exact character of this intermediate state depends on the magnitude of the orienting forces relative to that of the rotary Brownian movement. The intensity of this movement may be characterized in terms of a rotary diffusion constant, defined by equations exactly analogous to those defining the translational diffusion constant.

Inversely proportional to the rotary diffusion constant is the relaxation time, which is a measure of the time required for the molecules to revert to a random distribution, when an orienting force has been applied to the system and is then suddenly removed. Rotary diffusion constants and relaxation times are a function of molecular size and shape, and of the temperature and viscosity of the solvent medium. It is now necessary to define these quantities more precisely, before proceeding with a discussion of the methods by which they are determined.

### Rotary Diffusion Constants and Relaxation Times

The rotary diffusion constant may be defined by equations exactly analogous to Fick's two laws of translational diffusion. Consider a system of like molecules, in dilute solution, the orientation of each molecule in space being specified by the position of a particular axis located within the molecule.\* It will simplify the preliminary treatment of this problem to deal with the orientation of particles in only two dimensions; that is, with a system in which the principal axes of all the molecules lie in a single plane. Then the orientation of any molecule is given by the angle,  $\phi$ , between the principal axis of the molecule and the  $X$  axis. If  $\Delta n$  be the number of molecules in the system, per unit volume, having orientations lying between the two angles  $\phi$  and  $\phi + \Delta\phi$ , we may characterize the distribution of orientations throughout the system by the function  $\rho(\phi)$ .

$$\rho(\phi) = \lim_{\Delta\phi \rightarrow 0} \frac{\Delta n}{\Delta\phi} \quad (1)$$

The function  $\rho$ , in the treatment of rotary diffusion, is the analogue of the concentration in a system involving translational diffusion. If the

<sup>1</sup> There is, of course, no such thing as a perfectly isotropic liquid; the degree of isotropy of a liquid is not independent of that of its immediate surroundings. In the case of a liquid in contact with a solid surface, the molecules of the liquid are of course influenced by the surface, and the degree of isotropy of the liquid is not independent of that of the surface. In liquid crystals, of course, intermolecular forces are so great and so anisotropic that a large degree of spontaneous molecular orientation is produced. The discussion given here, however, will be confined to liquids which are isotropic to a high degree of approximation. Furthermore, in a dilute protein solution, in water, the protein molecules are so far apart that they do not attract or repel each other, the orientation of any one protein molecule will not be affected by the orientation of the other protein molecules. How dilute a given protein solution must be to permit this assumption to be made is the basis for most of the theo-

molecules are randomly distributed  $\rho(\phi) = \text{constant}$  and  $\frac{d\rho}{d\phi} = 0$ . If the distribution is not random, the net number of molecules,  $dn$ , whose orientation shifts in the time  $dt$  across the angle  $\phi$  from lower to higher values, due to the Brownian movement, is

$$dn = -\Theta \frac{d\rho}{d\phi} dt \quad (2)$$

where  $\Theta$ , the rotary diffusion constant, is a measure of the mobility of the particle in its rotation about an axis perpendicular to the principal axis. If the particle is an ellipsoid of revolution, we shall denote the axis of revolution as the  $a$  axis, and the rotation of this axis is rotation about the  $b(=c)$  axis. The corresponding diffusion constant is denoted by  $\Theta_b$ .

For a given value of  $\phi$ , the rate at which the function is changing with time is

$$\frac{\partial \rho}{\partial t} = \Theta \frac{\partial^2 \rho}{\partial \phi^2} \quad (3)$$

Equations 2 and 3 are the exact analogs, for rotary diffusion, of Fick's two laws of translational diffusion (Chapter 18, equations 1 and 3)<sup>2</sup>;  $\Theta$ , however, has different physical dimensions from  $D$ , its dimensions being  $t^{-1}$ . In what follows, numerical values of  $\Theta$  will be expressed in  $\text{sec}^{-1}$ .

The rotary diffusion constant,  $\Theta$ , like the translational diffusion constant,  $D$ , may be expressed in terms of the change in position of the molecules due to their Brownian movement. Imagine all the solute molecules oriented by an external force, so that their principal axes are parallel, and suppose this orienting force suddenly removed, so that the subsequent motion of the molecules is due only to their Brownian movement. After the lapse of a short time,  $t$ , a given molecule will have shifted through an angle  $\phi$  relative to its original orientation. Then  $\Theta$  is directly proportional to the mean square value of  $\phi$  for all the molecules in the system.

$$\Theta = 1/2 \frac{\overline{\phi^2}}{t} \quad (4)$$

The derivation of equations 3 and 4 is exactly analogous to the derivation of the corresponding equations for translational diffusion (see Chapter 18, equations 3 and 22).

Like  $D$ ,  $\Theta$  is directly proportional to the thermal energy of the molecule; similarly, it is inversely proportional to an "inner frictional constant",  $\zeta$ , which is a function of the size and shape of the molecule and the viscosity of the medium in which it is immersed. It is a measure of the torque which

must be applied to cause the molecule to rotate with unit angular velocity. By reasoning analogous to that employed in deriving equations 11 and 24 of Chapter 18 it is found that

$$\Theta = \frac{kT}{\zeta} \quad (5)$$

In the general case of an ellipsoid with three semi-axis,  $a$ ,  $b$ , and  $c$ , of different lengths, three different inner frictional constants are required to characterize the resistance of the medium to rotation of the molecule about each of these axes. There are three corresponding rotary diffusion constants

$$\Theta_a = kT/\zeta_a; \quad \Theta_b = kT/\zeta_b; \quad \Theta_c = kT/\zeta_c \quad (6)$$

We shall confine our subsequent discussion to ellipsoids of revolution, for which

$$\zeta_b = \zeta_c.$$

For many purposes, it is more convenient to characterize the rotary Brownian movement by another quantity, the relaxation time  $\tau$ . We may imagine the molecules oriented by an external force so that their  $a$  axes are all parallel to the  $X$  axis (which is fixed in space). If this force is suddenly removed, the Brownian movement leads to their disorientation. The position of any molecule after an interval of time may be characterized by the cosine of the angle  $\vartheta$  between its  $a$  axis and the  $X$  axis. (The molecule is now considered to be free to turn in any direction in space—its motion is not confined to a single plane, but instead may have components about both the  $b$  and  $c$  axes.) When the mean value of cosine  $\vartheta$  for the entire system of molecules has fallen to  $1/e$  ( $e = 2.718 \dots$  is the base of natural logarithms), the time that has elapsed is defined as the relaxation time,  $\tau_a$ , for motion of the  $a$  axis<sup>3</sup>. The relaxation time is greater, the greater the resistance of the medium to rotation of the molecule about this axis, and it is found that a simple reciprocal relation exists between the three relaxation times  $\tau_a$ ,  $\tau_b$  and  $\tau_c$ , for rotation of each of the axes, and the corresponding rotary diffusion constants defined in equation 6.

$$\tau_a = \frac{1}{\Theta_b + \Theta_c}; \quad \tau_b = \frac{1}{\Theta_a + \Theta_c}; \quad \tau_c = \frac{1}{\Theta_a + \Theta_b} \quad (7)$$

For an ellipsoid of revolution, these equations become:

$$\tau_a = \frac{1}{2\Theta_b}; \quad \tau_b = \tau_c = \frac{1}{\Theta_a + \Theta_b} \quad (7a)$$

**Relation to Molecular Size and Shape.** If the solute molecule is spherical, and is very large in comparison to the solvent molecules, the inner frictional constant  $\zeta$  is given by a formula due to Stokes

$$\zeta \text{ (sphere)} = 8\pi\eta r^3 \quad (8)$$

where  $r$  is the radius of the sphere and  $\eta$  is the viscosity of the solvent. The molecule is characterized by a single rotary diffusion constant  $\Theta$ , and a single relaxation time  $\tau$ .

$$\Theta = \frac{1}{2\tau} = \frac{kT}{8\pi\eta r^3} \quad (9)$$

This formula, due to Einstein, was experimentally verified by Jean Perrin<sup>4</sup>, by direct microscopic observation on spherical colloidal mastic particles (with radius  $6.5 \times 10^{-4}$  cm), which contained small enclosures of impurities on the surface, thus permitting their rotary motion to be directly followed.

Generalizing the hydrodynamic equations derived by Stokes for spheres, Edwardes<sup>5</sup> calculated the coefficients  $\zeta_1$ ,  $\zeta_2$  and  $\zeta_3$  for ellipsoids as a function of their axial ratios. The general equations are complicated; but for ellipsoids of revolution, which may be characterized by only two values of  $\zeta$ , they assume a simpler form, and have been employed by Gans<sup>6</sup> and Francois Perrin<sup>7</sup> to evaluate the rotary diffusion constants of molecules which may be treated as ellipsoids of revolution. The formulas of Gans and of Perrin are not identical, but the numerical values of  $\Theta$  calculated from them are nearly so, so that the formulas of either author may be used in practice. In the following discussion we shall employ Perrin's equations.

The volume of the ellipsoid (of axial ratio  $a/b$ ) is

$$V = \frac{4\pi}{3} ab^2 \quad (10)$$

The rotary diffusion constant,  $\Theta_0$ , and relaxation time,  $\tau_0$ , of a sphere of the same volume would be given by the relations

$$\Theta_0 = \frac{1}{2\tau_0} = \frac{kT}{8\pi ab^2\eta} \quad (11)$$

Consider first the case of an elongated ellipsoid of revolution ( $a > b$ ). Rotary Brownian movement of the  $a$  axis about the  $b$  axis is characterized by the relaxation time  $\tau_a$ , and the corresponding rotation diffusion constant  $\Theta_a = 1/2\tau_a$  (see equation 7a). These constants are conveniently ex-

<sup>4</sup> Perrin, J., *Compt. rend. Acad. Sci.*, 149, 549 (1909).

pressed by their values relative to those for a sphere of the same volume. Denoting by  $q$  the ratio  $b/a$ , Perrin's equation reads

$$\tau_a/\tau_0 = \frac{\Theta_0}{\Theta_b} = \frac{2(1-q^4)}{\frac{3q^2(2-q^2)}{\sqrt{1-q^2}} \ln \frac{1+\sqrt{1-q^2}}{q} - 3q^2} \quad (12)$$

$\tau_a/\tau_0$  is always greater than unity, and increases very rapidly as  $a/b$  increases. If  $a \gg b$ , equation 13 reduces approximately to the simpler formula:

$$\frac{\Theta_b}{\Theta_0} = \frac{\tau_0}{\tau_a} = \frac{3b^2}{2a^2} \left( -1 + 2 \ln \frac{2a}{b} \right) \quad (12a)$$

If  $a > 5b$ , the values of  $\Theta$  and  $\tau$ , calculated from formula 12a, agree within 1% with those calculated from the exact formula 12. Hence  $\Theta_b$ , from (11) and (12a), equals approximately:

$$\Theta_b = \frac{3kT}{16\pi\eta a^3} \left( -1 + 2 \ln \frac{2a}{b} \right) \quad (13)$$

Thus for a given value of  $a/b$ , the rotary diffusion constant of an elongated ellipsoid is inversely proportional to the cube of the length.

For the relaxation time,  $\tau_b$ , of an elongated ellipsoid of revolution involving rotation of the  $b$  axis, which may involve turning about both the  $a$  and the  $c(=b)$  axis, Perrin finds

$$\tau_b/\tau_0 = 2\Theta_0/(\Theta_a + \Theta_b) = \frac{4(1-q^4)}{\frac{3q^2(2q^2-1)}{\sqrt{1-q^2}} \ln \frac{1+\sqrt{1-q^2}}{q} + 3} \quad (14)$$

If  $a \gg b$ , ( $q \rightarrow 0$ ), this reduces approximately to

$$\tau_b/\tau_0 = 2\Theta_0/(\Theta_a + \Theta_b) = \frac{4}{3 \left( 1 - \frac{b^2}{a^2} \ln \frac{2a}{b} \right)} \quad (14a)$$

Thus  $\tau_b$  for an elongated ellipsoid is always greater than  $\tau_0$ , but never becomes greater than  $4\tau_0/3$ , even when  $a/b$  becomes infinite.

For a flattened ellipsoid ( $a < b$ , and  $q > 1$ ) a different set of equations holds. For rotation about the equatorial ( $b$ ) axis

$$\tau_a/\tau_0 = \Theta_0/\Theta_b = \frac{2(1-q^4)}{\frac{3q^2(2-q^2)}{\sqrt{q^2-1}} \tan^{-1} \sqrt{q^2-1} - 3q^2} \quad (15)$$

and for rotation about the  $a$  axis (axis of revolution)

$$4(1-q^4)$$



If  $b \gg a$ , both equations 15 and 16 reduce approximately to

$$\tau_a/\tau_0 \cong \tau_b/\tau_0 \cong (2b/a)/[3 \tan^{-1}(b/a)] \cong 4b/3\pi a \quad (16a)$$

and substituting the value of  $\tau_0$  from 11

$$\tau_a \cong \tau_b \cong 16\eta b^3/3kT \cong 1/2\theta_b \cong 1/2\theta_a \quad (16b)$$

Thus for a flattened ellipsoid, as  $b/a$  becomes infinite, both relaxation times become infinite, and the corresponding rotary diffusion constants approach zero. When  $b/a$  is very large, the relaxation times are proportional to the cube of the  $b$  semi-axis, and do not depend at all on the length of the  $a$  semi-axis.

### The Viscosity of Protein Solutions

The viscosity of a liquid is a measure of its resistance to shearing stress. The viscosity coefficient is the force required, per unit area, to maintain unit difference of velocity between parallel planes in the liquid, when the distance between the planes is unity. Thus the tangential force,  $F$ , required to maintain the relative velocity  $\Delta v$  between two parallel planes separated by a distance  $\Delta y$ , in the direction normal to the plane of flow, is

$$F = A\eta \frac{\Delta v}{\Delta y} \quad (17)$$

where  $A$  is the area of one of the planes of flow. The viscosity coefficient,  $\eta$ , of the liquid is defined by this equation. If  $\Delta y$  approaches zero, we may write it in the differential form

$$F = A\eta \frac{dv}{dy} \quad (18)$$

$\frac{dv}{dy}$  is known as the velocity gradient. The dimensions of the viscosity coefficient,  $\eta$ , are  $ml^{-1}t^{-1}$ . The unit of viscosity, expressed in cgs. units, is known as the poise. The viscosity of water at  $20^\circ$  is almost exactly 0.01 poise.

Equations 17 and 18 were first derived by Newton, who made the assumption that the viscosity coefficient is independent of the velocity gradient in the flowing liquid. The consequences of this assumption have been very exactly verified for most pure liquids over a very wide range of velocity gradients. The flow of such liquids is said to be Newtonian. There are, however, important classes of liquids, especially certain colloidal solutions, for which this assumption is invalid. In such cases, experiments almost invariably indicate that  $\eta$  decreases with increasing

**Measurement of Viscosity.** There are two methods of major importance for measuring the viscosity of liquids<sup>8</sup>. The first of these depends on the rate of flow of the liquid through a capillary tube of known radius and length, under a known pressure. If the rate of flow is not too rapid<sup>9</sup>, the liquid flows through the capillary in concentric cylindrical laminae, the velocity being uniform throughout any lamina. It is generally assumed—and the consequences of the assumption agree with experiment—that the layer of liquid immediately adjacent to the wall of the capillary adheres to the wall and does not move. A straightforward calculation from dynamical principles then gives for the velocity of the liquid<sup>10</sup>

$$v = \frac{\Delta p}{4\eta l}(a^2 - r^2) \quad (19)$$

where  $\Delta p$  is the pressure gradient acting on the liquid,  $l$  is the length of the capillary,  $a$  its radius, and  $r$  the distance from the center of the capillary. The velocity is maximal at the center of the capillary ( $r = 0$ ), zero at the wall ( $r = a$ ). On the other hand, the velocity gradient

$$\frac{dv}{dr} = \frac{-\Delta p \cdot r}{2\eta l} \quad (20)$$

is zero at the center of the capillary, and increases linearly with the distance from the center. The volume of liquid passing any cross-section of the capillary in unit time is

$$Q = 2\pi \int_0^a vr \, dr = \frac{\pi \cdot \Delta p \cdot a^4}{8\eta l} \quad (21)$$

Thus, if  $\Delta p$ ,  $l$ ,  $a$  and  $Q$  are all measured, the viscosity coefficient may be calculated. If it is sufficient to know the viscosity of the liquid relative to that of a known standard liquid,  $Q$  and  $\Delta p$  may be determined for both the standard and the unknown liquid; and an explicit determination of  $l$  and  $a$  is then unnecessary.

The other type of viscometer most commonly used was first developed in detail by Couette, and is commonly called after him. It consists essentially of two vertical coaxial cylinders, with the liquid in the gap between them. The outer cylinder is rotated at constant speed; it thereby impresses a rotatory motion on the adjoining layer of liquid, which is in turn transmitted to the next layer, and so forth. When the system has attained a

<sup>8</sup> A third method, based on the measurement of the rate of flow of a liquid through a capillary tube of known radius and density, is described in D. van Nostrand, "The Viscosity of Liquids,"

<sup>9</sup> Above a certain critical velocity of flow, the magnitude of which depends on the viscosity and density of

steady state, a turning moment is exerted on the fixed inner cylinder. This moment is calculated<sup>8, 10</sup> to be

$$M = 4\pi L\eta\Omega R_1^2 R_2^2 / (R_2^2 - R_1^2) = K\eta\Omega \quad (22)$$

where  $L$  is the length of the cylinders,  $R_1$  the radius of the inner cylinder,  $R_2$  that of the outer cylinder, and  $\Omega$  is the angular velocity of the outer cylinder.  $K$  is thus a constant of the apparatus.  $M$  is generally measured by suspending the inner cylinder from a wire of known restoring force,  $N$ , and measuring the angle  $\zeta$  through which it is deflected. We then have:

$$\zeta N = M = K\eta\Omega \quad (23)$$

or, since  $N$  is also a constant for a given wire,

$$\zeta = K'\eta\Omega \quad (24)$$

For a given liquid at a given temperature  $\zeta$  is directly proportional to  $\Omega$ . If the deflections  $\zeta$  and  $\zeta'$  are found for two different liquids at the angular velocities  $\Omega$  and  $\Omega'$ , the ratio of their viscosities is

$$\frac{\eta'}{\eta} = \frac{\zeta'\Omega}{\zeta\Omega'} \quad (25)$$

The angular velocity,  $\omega$ , of the liquid at any radius,  $r$ , ( $R_1 \leq r \leq R_2$ ) is found to be

$$\omega = \left( \frac{1/R_1^2 - 1/r^2}{1/R_1^2 - 1/R_2^2} \right) \Omega \quad (26)$$

and the velocity gradient is

$$\beta = r \frac{d\omega}{dr} = 2\Omega \left( \frac{1/r^2}{1/R_1^2 - 1/R_2^2} \right) \quad (27)$$

If the gap between the cylinders is very small [ $(R_2 - R_1) \ll R_1$ ] the velocity gradient is approximately constant and equal to

$$\beta \cong \frac{\Omega R_1}{R_2 - R_1} \cong \frac{\Omega R_2}{R_2 - R_1} \quad (28)$$

**The Viscosity of Solutions of Large Molecules.** We are not directly concerned in this discussion with the vast number of studies that have been carried out on the viscosity of pure liquids, and on solutions containing only small molecules<sup>11</sup>. Our concern here is with protein molecules, which are very large in comparison with the molecules of the solvent, so that the latter may be regarded approximately as a continuum, characterized by its macroscopic viscosity,  $\eta_0$ . If very large solute molecules

are introduced into the solvent, it is invariably found that the viscosity of the solution is greater than that of the pure solvent. This effect may be qualitatively understood in terms of the energy required to maintain viscous flow. In a pure liquid, in which a uniform velocity gradient  $\beta$  is maintained throughout the system<sup>12</sup> the expenditure of energy per unit time necessary to maintain viscous flow in a unit volume of the liquid is<sup>13</sup>

$$\frac{1}{\text{Vol.}} \frac{dW}{dt} = \eta \beta^2 \quad (29)$$

The introduction of very large solute molecules involves additional work in maintaining the flow, because of the distortion of the stream lines by the large molecules. Therefore the system as a whole behaves like a pure liquid of higher viscosity.

The first quantitative calculation of the effect of large solute molecules on the viscosity of a pure liquid was made by Einstein<sup>14</sup>. He treated the special case of spherical, rigid, uncharged solute molecules, assuming no slip between solvent and solute (that is, adherence of a layer of solvent to the solute molecules). By straightforward hydrodynamic reasoning, he obtained, for the viscosity,  $\eta$ , of the solution

$$(\eta/\eta_0) - 1 = 2.5\Phi \quad (30)$$

where  $\Phi$  is the volume fraction of the system occupied by the solute molecules. It is apparent from this equation that the volume of the individual spheres has no effect on the result; only the total volume occupied by the spheres in the system is of importance. This equation is valid only as long as  $\Phi$  is small ( $<0.03$ , approximately). It has been experimentally confirmed by Eirich, Bunzl and Margaretha<sup>15</sup>, for suspensions of yeast, mastic, paraffin, sulfur and various approximately spherical spores and bacteria. These extensive investigations, together with other more isolated studies, appear to demonstrate the validity of Einstein's equation for systems of the class to which it was intended to apply.

It will be found convenient to write the quantity

$$\frac{1}{\Phi} \left( \frac{\eta}{\eta_0} - 1 \right) \equiv \nu \quad (31)$$

The quantity  $\nu$  may be called the "viscosity increment" of the solute<sup>15a</sup>.

<sup>12</sup> This condition is realized in practice in a Couette viscometer under the conditions described by equation 28.

<sup>13</sup> See for instance J. R. Robinson, *Proc. Roy. Soc. London*, A170, 519 (1939).

<sup>14</sup> See for instance J. R. Robinson, *Proc. Roy. Soc. London*, A170, 519 (1939).

Thus Einstein's equation for spheres may be written

$$\nu = 2.5 \quad (30a)$$

if  $\Phi < 0.03$ .

It is found that for non-spherical solute particles,  $\nu$  is always greater than 2.5. The data in this field are very extensive, largely owing to the work of Staudinger and his collaborators<sup>16</sup> on solutions of long chain polymers of different degrees of polymerization. Staudinger found an approximate proportionality between  $\nu$  and molecular weight in many such series (polystyrol, polyoxymethylene, and many other similar polymers)

$$\nu = K_m M \quad (32)$$

where  $K_m$  is a constant, and  $M$  is the number of monomer units in the polymer molecule<sup>17</sup>. These observations are best explained by the assumption that the polymer molecules, owing to free rotation, are bent and kinked in a nearly random fashion in solution, and that the actual mean length from end to end of the polymer molecule is proportional to  $\sqrt{M}$ <sup>18</sup>, a relation which may be derived from statistical reasoning. For proteins, also, the values of  $\nu$  found experimentally (Table 2) are always greater than 2.5, and often very much greater. It has frequently been suggested that these values may be explained by hydration of the protein. If one gram of protein combines with  $w$  grams of solvent, so that the molecules so bound do not take part in the motion of the liquid, then the volume fraction  $C_h$  of the hydrated protein is related to the volume fraction  $C_a$  of the anhydrous protein by the relation

$$C_h = C_a \left( 1 + \frac{w}{V_a \rho} \right) = H C_a \quad (33)$$

provided  $C_h \ll 1$ .

Here  $V_a$  is the partial specific volume of the anhydrous protein,  $\rho$  is the density of the solvent, and  $H$  is a "solvation factor" defined by equation 33<sup>19</sup>. The determination of  $w$  in protein crystals has been described in Chapter 16 (see also Chapter 14, Table 1). The highest reliable value

$$[\eta] = \lim_{g \rightarrow 0} \frac{1}{g} \left( \frac{\eta}{\eta_0} - 1 \right) \quad (31a)$$

Here  $g$  is the concentration of solute in grams per 100 cc of solution. Hence

$$\frac{[\eta]}{\nu} = \frac{V}{100} \quad (31b)$$

where  $V$  is the partial specific volume of the solute. This is the same as the definition of  $V$  in Staudinger, H., "Die Hochmolekularen Organischen Verbindungen", Berlin (1932); *Helv. Chim. Acta*, 19, 204 (1936).

is that of McMeekin and Warner for  $\beta$ -lactoglobulin, which is approximately 0.8. There is good reason to believe, however, that hydration of protein molecules in solution is lower than in crystals. Much of the water in wet protein crystals occupies large interstices between the protein molecules, and would not be expected to remain fixed to these molecules when they pass into solution in water or other solvents. The water packed by electrostriction around the charged groups is certainly quite tightly held, but the value of  $w$  due to this factor alone would be very small for most proteins—in the neighborhood of 0.03, as estimated by the calculations given in Chapter 16. Probably some water is fairly strongly bound to other polar groups in the protein, although less firmly held than the water of electrostriction; but no exact estimate of its amount can be given. For purposes of calculation we may consider the consequences of assuming a value of  $w$  equal to 0.5, although this is almost certainly higher than the values to be expected of typical proteins. Setting  $w = 0.5$ ,  $V_a = 0.75$ , and  $\rho = 1$  in equation 33, we find  $H = 1.67$ . Thus, if the protein is spherical,  $\nu$  in equation 31 is 2.5 if  $C$  is expressed as  $C_h$  (equation 33), and the value of  $\nu$  expressed in terms of the volume fraction of *anhydrous* protein is equal to  $2.5H$ , or approximately 4.2. This is probably close to the upper limit of the values of  $\nu$  explicable solely in terms of hydration. The values actually found for many proteins, however, are much higher than this, and can be explained in terms of deviation of the protein molecule from the spherical shape.

#### Rotary Motion of Ellipsoidal Particles Subjected to a Velocity Gradient.

The first theoretical treatment of this problem was given by Jeffery<sup>20</sup> who extended Einstein's calculations to the more complicated problem of large ellipsoids immersed in a viscous fluid. In this case, the effect of the ellipsoid in disturbing the streaming of the fluid, and thereby increasing the viscosity, depends upon the orientation relative to the stream lines, and this orientation is constantly changing because of the Brownian movement of the ellipsoid, and because of the rotation produced by the velocity gradient in the liquid. Jeffery considers only the case where the Brownian movement is negligible, and the orientation of the ellipsoids is determined by hydrodynamic forces. The orientation of the molecule at any moment in a streaming liquid of constant velocity gradient is shown in Fig. 1<sup>21</sup>; it is described in terms of the angles  $\theta$  and  $\phi$ .  $\theta$  is the angle between the  $a$  axis of the ellipsoid and the positive  $Z$  axis ( $\theta = 90^\circ$  if the  $a$  axis lies in the plane of the paper);  $\phi$  is the angle between the  $XZ$  plane, and the plane

<sup>20</sup> Jeffery, G. B., *Proc. Roy. Soc. London*, **A102**, 161 (1922-23).

passing through the  $Z$  axis of coördinates and the  $a$  axis of the ellipsoid. If  $a$  is the magnitude of the semi-axis of revolution of the ellipsoid and  $b$  that of the equatorial semi-axis, Jeffery finds for the components of the rotation of the ellipsoid due to the velocity gradient,  $\beta$ :

$$\frac{d\phi}{dt} = \frac{-\beta(a^2 \sin^2 \phi + b^2 \cos^2 \phi)}{a^2 + b^2} \quad (34)$$

$$\frac{d\theta}{dt} = \beta \frac{(a^2 - b^2)}{a^2 + b^2} \sin \theta \cos \theta \sin \phi \cos \phi \quad (35)$$

The effect of the velocity gradient in the liquid is to cause a continued rotation of the ellipsoid. If the ellipsoid is prolate ( $a > b$ ) it rotates most rapidly ( $d\phi/dt$  maximum) when the  $a$  axis is perpendicular to the stream lines ( $\phi = 90^\circ$ ), most slowly when this axis is parallel to the stream lines

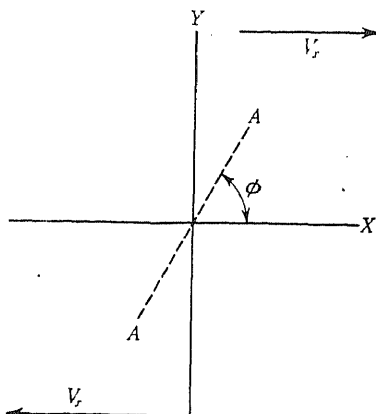


Figure 1.

Orientation of an ellipsoidal molecule in a flowing liquid of constant velocity  $V_x$ . The positive  $Z$  axis points upward from the plane of the paper. The projection of the axis of revolution of the ellipsoid on the  $XY$  plane is denoted by  $AA$ . The movement of the liquid is parallel to the  $X$  axis, and is described by the equation  $V_x = GY$ , where  $V_x$  is the velocity and  $G$  is the velocity gradient. The significance of the angle  $\phi$  is shown in the figure ( $\phi = 0$  when the  $a$  axis of the ellipsoid lies in the  $XZ$  plane).  $\theta$  is the smaller of the two angles between the  $a$  axis of the ellipsoid and the positive  $Z$  axis. The origin is taken at the center of the ellipsoid.

( $\phi = 0$ ). Thus a very elongated ellipsoid ( $a \gg b$ ) spends<sup>22</sup> most of its time, with the  $a$  axis parallel or nearly parallel to the stream lines. Conversely, a disc-shaped ellipsoid ( $a \ll b$ ) spends most of its time with the  $b$  axis parallel to the stream lines.

The contribution of a single ellipsoid to the viscosity of the system is calculated by Jeffery for any given orientation; to calculate the viscosity of the system as a whole, it is necessary to average over all orientations, allowing for the relative time spent by the particles in each orientation. Jeffery was unable to reach a unique solution of this problem, because the distribution of particles in various orientations, with respect to the angle  $\theta$ ,

ment. He did, however, give maximum and minimum values for  $\nu$ , as a function of the ratio of the axes of the ellipsoids.

**Relation of Viscosity Increment to Shape of Molecules.** Recently a thorough treatment has been given by Simha<sup>23</sup> of the viscosity of solutions of ellipsoidal molecules, in which the Brownian movement is so intense that the orientation produced by the shearing forces is negligible. This is true provided that the ratio  $\alpha$  of the velocity gradient to the rotary diffusion constant of the molecules is much less than unity. All orientations then are equally probable, and the motion of the ellipsoids with respect to the angle  $\phi$  is independent of their motion with respect to  $\theta$ . Simha then calculates  $\nu$  for ellipsoids arranged at random, employing the formula de-

TABLE 1

Axial Ratio	elongated $\nu$	flattened	Axial Ratio	elongated	flattened
1.0	2.50	2.50	20.0	38.6	14.80
1.5	2.63	2.62	25.0	55.2	18.19
2.0	2.91	2.85	30.0	74.5	21.6
3.0	3.68	3.43	40.0	120.8	28.3
4.0	4.66	4.06	50.0	176.5	35.0
5.0	5.81	4.71	60.0	242.0	41.7
6.0	7.10	5.36	80.0	400.0	55.1
8.0	10.10	6.70	100.0	593.0	68.6
10.0	13.63	8.04	150.0	1222.0	102.3
12.0	17.76	9.39	200.0	2051.0	136.2
15.0	24.8	11.42	300.0	4278.0	204.1

From Mehl, J. W., Oncley, J. L., and Simha, R., *Science*, **92**, 132 (1940).

rived by Jeffery describing the increase in viscosity produced by a molecule in any given orientation. The calculations give (if  $J = a/b$ )

$$\text{For rods: } \nu = \frac{J^2}{15(\log 2J - 3/2)} + \frac{J^2}{5(\log 2J - 1/2)} + (14/15) \quad (36)$$

$$\text{For discs: } \nu = (16/15)(J/\tan^{-1}J) \quad (37)$$

The first term on the right of equation 36 had already been derived, in substantially the same form, by Eisenschitz<sup>24</sup> and Burgers<sup>25</sup>; the other terms are derived by Simha's more rigorous treatment.

The numerical relations between  $\nu$  and the axial ratio of the ellipsoid ( $J$  or  $1/J$ ) are given for convenience in Table 1. With the aid of these tables, and the values of  $\nu$  obtained for a number of proteins by Polson<sup>26</sup> and other authors, Mehl, Oncley and Simha<sup>27</sup> calculated the shape factors for a variety of proteins. These are listed in Table 2, along with similar factors derived from measurements of  $f/f_0$  determined from sedimentation

<sup>23</sup> Simha, R., *J. Physical Chem.*, **44**, 25 (1940).



The values of  $f/f_0$  for the protein studied were assumed to be subject to an error of  $\pm 4\%$ , and those of  $\nu$  to an error of  $\pm 10\%$ . The areas included between pairs of such curves would then include the possible choices for shape and hydration. Such curves for egg albumin and thyroglobulin

Protein	$f/f_0$	$\nu$	a/b, elongated		b/a, flattened	
			Diffusion	Viscosity	Diffusion	Viscosity
Egg albumin.....	1.17	5.7	3.8	5.0	4.0	6.7
Serum albumin.....	1.25	6.5	5.0	5.6	5.4	7.7
Hemoglobin.....	1.16	5.3	3.7	4.6	3.9	6.0
Amadin.....	1.28	7.0	5.4	6.0	6.0	8.5
Octopus hemocyanin.....	1.38	9.0	7.2	7.3	8.2	11.4
Gliadin.....	1.60	14.6	10.9	10.5	13.6	21
Homarus hemocyanin.....	1.27	6.4	5.2	5.5	5.8	7.5
Helix pom. hemocyanin.....	1.24	6.4	4.8	5.5	5.2	7.5
Serum globulin.....	1.41	9.0	7.6	7.3	8.9	11.4
Thyroglobulin.....	1.43	9.9	7.8	7.9	9.2	12.7
Lactoglobulin.....	1.26	6.0	5.2	5.1	5.7	6.9
Pepsin.....	1.08	5.2	2.5	4.5	2.6	5.8
Helix hemocyanin pH 8.6.....	1.89	18.0	16.6	12.0	23.9	26

are shown in Fig. 2b; they were chosen as well characterized examples of fairly symmetrical and fairly asymmetrical molecules. The shape of these curves is such that they never cross sharply, but it appears that for these two cases hydration greater than 0.5 gram of water per gram of protein, and the flattened shape, appear relatively improbable. Even assuming a slightly greater degree of hydration, most of the proteins in Table 2 must deviate appreciably from the spherical shape; the only ones which might be approximately spherical are egg albumin, hemoglobin, pepsin and possibly lactoglobulin<sup>28</sup>. Recently Neurath and Saum<sup>29</sup> have investigated the viscosity increment and the diffusion constant of serum albumin in water

and in urea solutions of varying concentration. The work of Burk<sup>30</sup> has shown that the molecular weight of this protein is essentially the same in concentrated urea solution as in water, but its diffusion constant progressively decreases, and its viscosity increment,  $\nu$ , steadily increases, with increasing concentration of urea (Table 3). The results can best be interpreted as revealing a progressive unfolding of the protein, which in water is not far from spherical, with increasing concentration of urea. This

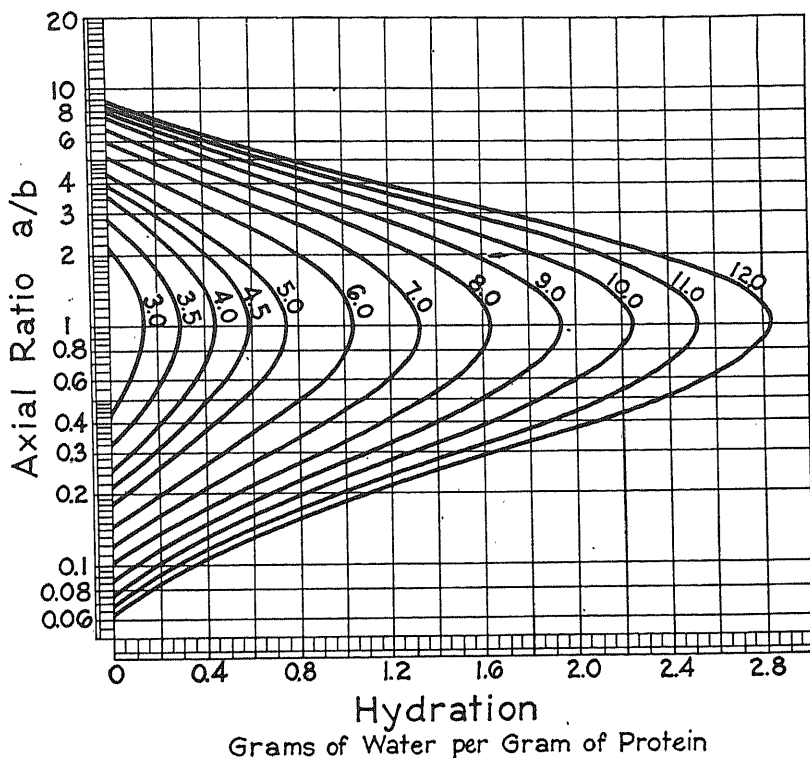


Figure 2a. Values of axial ratio and hydration in accord with various viscosity coefficients. (Contour lines denote  $\nu$  values) From Oncley, J. L., *Annals New York Acad. Sci.*, **41**, 121 (1941).

interpretation is in harmony with the results of x-ray diffraction studies on denatured proteins. The estimated shape factors from viscosity and from diffusion measurements agree reasonably well. The increase in the viscosity of egg albumin produced by urea and heat denaturation has been studied by Bull<sup>31</sup>

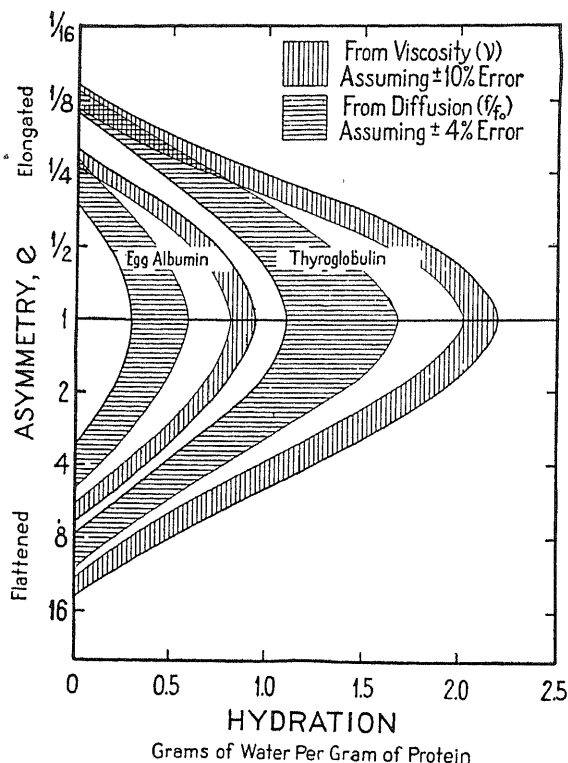


Figure 2b. The asymmetry of protein molecules calculated from viscosity and diffusion, assuming various degrees of hydration. From Mehl, J. W., Oncley, J. L., and Simha, R., *Science*, **92**, 132 (1940).

TABLE 3. EFFECT OF UREA ON THE VISCOSITY INCREMENT AND DIFFUSION CONSTANT OF SERUM ALBUMIN

Urea moles/l.	Viscosity increment $\nu$	$a/b$ , calc. from Eq. 36	$D' = \text{corrected}$ diffusion const. $\times 10^7$	$D_0/D'$	$a/b$ from diffusion
0.0	6.50	5.6	6.85	1.20	4.3
0.5	6.60	5.7	6.20	1.33	6.1
1.5	7.40	6.2	6.08	1.35	6.5
3.0	8.7	7.2	5.69	1.44	8.0
4.5	16.4	11.4	4.45	1.85	16.5
6.0	19.6	12.9	4.27	1.93	18.3
6.66	22.6	14	4.15	1.98	19.4

The corrected diffusion constant  $D' = C(\eta/\eta_0)$ , where  $D$  is the observed diffusion constant, and  $\eta/\eta_0$  is the viscosity of the solvent ( $\eta$ ) relative to that of water ( $\eta_0$ ). Values of  $D$  and  $D'$  in  $\text{cm}^2 \text{sec}^{-1}$ .  $D_0$  is the calculated diffusion constant for a spherical molecule of mol. wt. 67,100 =  $8.22 \times 10^{-7} \text{ cm}^2 \text{sec}^{-1}$  at  $25^\circ$  in water. Values in the last column are obtained from the  $D_0/D'$  values by assuming  $a/b = 4.3$  at 0.0 M urea.

**Electroviscous Effects.** It might well be supposed that variations of pH would affect the observed values of  $\nu$ . Laqueur and Sackur<sup>31a</sup> nearly forty years ago discovered that the viscosity of casein was markedly affected by the addition of acid or alkali. The same effect was also clearly shown in the work of Hardy<sup>31b</sup> on serum globulin. Hardy showed that the viscosity of serum globulin is generally a minimum near the isoelectric point and rises to a considerably higher value on the addition of either acid or alkali. Similar effects in gelatin and other proteins were intensively studied by Loeb<sup>31c</sup>. Loeb and others also showed that the rise in viscosity produced by acids or alkalis can be partially or wholly abolished by the addition of neutral salts<sup>32</sup>.

It appears probable that the change in the viscosity of a protein solution on the addition of acid or alkali is due predominantly to electroviscous effects and can be explained by the drag exerted by the charged protein on the liquid in the surrounding electrical double layer. On the basis of these considerations, Smoluchowski<sup>33</sup> proposed a formula for spherical particles of radius  $r$ , which in our notation may be written:

$$\nu_{\text{sphere}} = 2.5 \left( 1 + \frac{D\psi_0}{2\pi r^2 x \eta_0} \right) \quad (38)$$

Here  $D$  = dielectric constant of solvent;  $\psi_0$  = electrokinetic potential; and  $x$  = specific conductivity of the solution. Addition of salts to the solution can reduce the second term in 38 practically to zero, and give  $\nu$  the theoretical value for an uncharged particle. It was indeed found by Polson<sup>26</sup> that the viscosity of egg albumin solutions is independent of pH between pH 4.56 and 6.67, provided the measurements are made in a solution containing 0.20M NaCl. It is very important therefore, when measuring the viscosity of protein solutions in order to calculate their shapes, to work in solutions of fairly high ionic strengths, and at several different pH values.

**Change of Viscosity with Velocity Gradient: "Non-Newtonian" Flow.** In the preceding treatment it has been assumed that the axes of the protein molecules are distributed at random in the following solution. This is true if  $\alpha = \beta/\theta \ll 1$ . This condition is certainly fulfilled for most of the proteins listed in Table 2. Values which have been found for  $\theta$ , the rotary diffusion constant, vary from  $6 \times 10^6 \text{ sec}^{-1}$  for hemoglobin to  $3 \times 10^6 \text{ sec}^{-1}$  for serum pseudoglobulin<sup>34</sup>. The mean velocity gradients

<sup>31a</sup> Laqueur, E., and Sackur, O., *Hofmeister's Beiträge*, 3, 193 (1903).

<sup>31b</sup> Hardy, S. A., *Journal of Biological Chemistry*, 33, 251 (1905).

<sup>31c</sup> Loeb, J., *Theory of Colloids*, McGraw Hill Book Co., New York, 1923.

<sup>32</sup> Loeb, J., and others, *Journal of Biological Chemistry*, 19, 1, 1906. <sup>33</sup> Smoluchowski, M., *Ann. Physik*, 17, 1, 1906. <sup>34</sup> Valkó, E., *Kolloidchemie der Eiweißkörper*, 1923.

employed<sup>35</sup> in most capillary viscometers vary from a few hundred to a few thousand  $\text{sec}^{-1}$ ; in viscometers of the Couette type, the gradients are generally much lower than this. In such systems, therefore,  $\beta/\theta < 0.01$ . The situation is entirely different, however, when very elongated protein molecules, such as myosin, collagen and tobacco mosaic virus are considered. These are all exceedingly elongated molecules; as shown by measurements of double refraction of flow, their length is of the order of magnitude of several thousand Ångströms, and their rotary diffusion constants are near  $10 \text{ sec}^{-1}$  (or less). Such molecules are very highly oriented even at moderate velocity gradients. This means that the measured viscosity is a function of the velocity gradient. At very low gradients, the rod-shaped molecules are distributed at random, and their resistance to the viscous flow of the liquid is very great. Correspondingly, the viscosity increment under these conditions—which we may denote by the symbol  $\nu_0$ —is high<sup>36</sup>. At higher gradients ( $\alpha \cong 1$ ) the molecules deviate appreciably from the state of purely random distribution; and at very high velocity gradients ( $\alpha \gg 1$ ) they are oriented predominantly parallel to the stream lines. In this position their influence in increasing the viscosity of the solution is a minimum, and  $\nu$  tends toward a lower limiting value for “infinite velocity gradient,” which we may denote by  $\nu_\infty$ . The more the ellipsoidal molecule deviates from the spherical shape, the greater will be the difference between  $\nu_0$  and  $\nu_\infty$ ; the smaller the rotary diffusion constant of the molecule, the lower will be the velocity gradient, at which  $\nu$  becomes effectively equal to  $\nu_\infty$ <sup>37</sup>. All these considerations have been treated in detail by Peterlin<sup>38</sup>, who has given a very elaborate mathematical formulation of the problem. Peterlin gives curves showing  $\nu$  as a function of  $\alpha$  for certain specific values of  $J$ , both for elongated and for flattened ellipsoids, but he did not derive a general formula for  $\nu_\infty$  as a function of  $J$ , on account of mathematical difficulties. There also remain certain unresolved problems regarding the underlying assumptions made by Peterlin, but the general picture given by his treatment would appear to be correct.

As an example we may compare the measurements of Robinson<sup>39</sup> and of Lauffer<sup>40</sup> on the viscosity of tobacco mosaic virus. Robinson, employing a viscometer of the Couette type, and velocity gradients varying from less than 1 to about  $10 \text{ sec}^{-1}$ , obtained a viscosity increment,  $\nu_0$ , for the virus of 1500 in a 0.02% solution, corresponding from Table 1 to an axial ratio of about 160 to 1. Lauffer, employing a capillary viscometer, in which the

<sup>35</sup> Kroepelin, H., *Kolloid-Z.*, **47**, 294 (1929).

<sup>36</sup> All the values of  $\nu$  previously discussed in this chapter are  $\nu_0$  values. Hitherto we have not explicitly introduced the subscript 0, since only conditions of random orientation were being considered.

mean velocity gradient was certainly far higher, found  $\nu = 55.8$  in a 0.0296% solution. Even this value may be somewhat higher than  $\nu_{\infty}$ . These measurements were made on different preparations of virus, and are therefore not strictly comparable. They illustrate clearly, however, the profound effect of the velocity gradient on the viscosity increment for highly asymmetrical proteins.

A detailed study of myosin in solutions of varying concentration of protein and velocity gradient was made by Edsall and Mehl<sup>41</sup>, using a

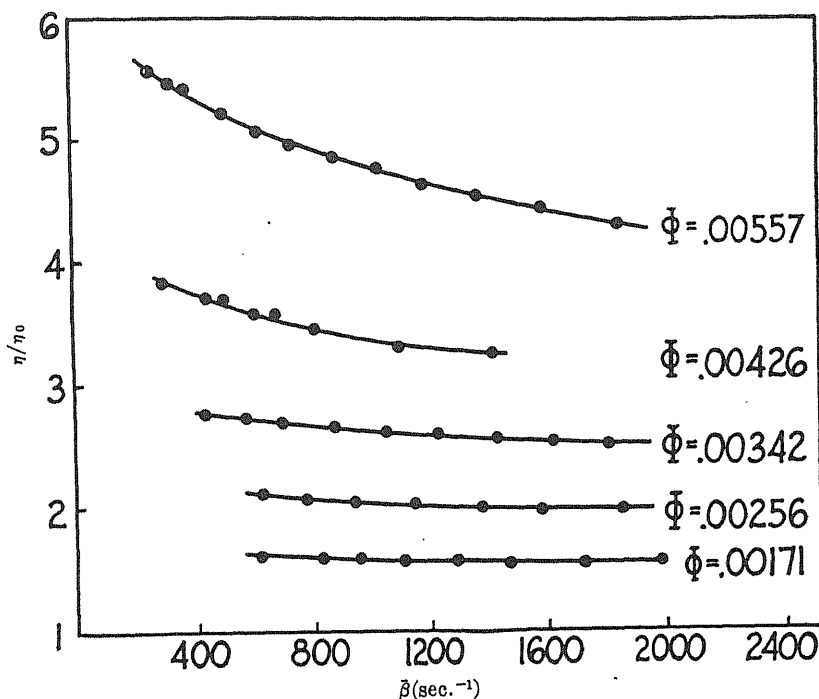


Figure 3 Relative viscosity of rabbit myosin solutions, plotted as a function of the mean velocity gradient. The volume fraction ( $\Phi$ ) of myosin in each solution studied is indicated to the right of each curve. Temperature 25°. From Edsall, J. T., and J. B. Mehl, *J. Biol. Chem.*, **133**, 409 (1940).

capillary viscometer. Their results are summarized in Fig. 3, in which the volume fraction ( $\Phi$ ) of protein is calculated from the weight fraction assuming the apparent specific volume = 0.75. The velocity gradient is a mean value,  $\bar{\beta}$ , for the flowing liquid in the capillary, calculated from the formula of Kroepelin<sup>35</sup>:

where  $V$  is the volume of liquid flowing through the capillary of radius  $r$  in time  $t$ . Kroepelin has shown that  $\nu$  is a uniquely defined function of  $\bar{\beta}$  for capillaries of varying radius. All of the mean velocity gradients shown in Fig. 3 are relatively high, and in all cases  $\nu$  falls with increasing  $\bar{\beta}$ , approaching a limiting value which may be taken as  $\nu_{\infty}$ . This value, however, is not a constant, but is a function of concentration, falling from approximately 530 in the most concentrated solution shown in Fig. 3, to 340 in the most dilute. If  $\nu_{\infty}$  is plotted against  $\Phi$ , an extrapolated value of  $\nu_{\infty} = 240$  is obtained for zero concentration of myosin. The higher values of  $\nu_{\infty}$  obtained at finite concentrations are due, not only to the interaction of the flowing solvent with the individual myosin molecules, but to the interaction of the myosin molecules with one another, probably involving the interference of one myosin molecule with the rotation of others. Such concentration effects must be sharply distinguished from the effects which remain operative even at infinite dilution of the solute. The latter are a function of the shape and hydration of the solute molecule. The effects of concentration are likely to be particularly apparent in solutions of very long molecules, which may interfere or become entangled during flow; such interference increases the measured viscosity, and also magnifies the change of viscosity with velocity gradient<sup>42</sup>. Neither of these concentration effects has yet been interpreted quantitatively in theoretical terms. It is therefore very important, if viscosity measurements are to be used in the determination of shape factors, to work at low protein concentrations. Since the viscosity increment  $\nu_0$  for random distribution is more readily interpreted than  $\nu_{\infty}$ , it is also desirable to work at low velocity gradients. This involves the use of the Couette viscometer, which has the further advantage that the velocity gradient can be quite precisely defined, and varies only over a relatively narrow range between the inner and the outer cylinders; whereas in the capillary viscometer, the gradient varies from zero at the centre of the capillary to a very high value at the wall. More general use of the Couette viscometer by workers in this field should greatly advance our understanding, particularly of very elongated protein molecules.

There can be no doubt that, at low velocity gradients, the  $\nu_0$  values for myosin should be very much higher than any of the  $\nu$  values reported by Edsall and Mehl for high velocity gradients. The  $\nu_{\infty}$  values, as reported by these authors, reveal, however, a very pronounced effect of various

<sup>42</sup> We have not here discuss  
for very low shearing stresses;  $\nu$

behave like rigid gels  
iscous fluids. Gelatin  
must be chitin

denaturing agents (Table 4). The effect of guanidine, urea and ammonium chloride is to *decrease* the viscosity of myosin solutions; and the same was found to be true from qualitative observations for a very large number of other denaturing agents. This is most readily explained on the assumption that the very elongated molecules of native myosin are broken up into shorter chains by guanidine, urea, and other agents. In harmony with this explanation is the effect of urea in decreasing the mean molecular weight of myosin in solution (determined by osmotic pressure) from about  $10^6$  to about  $10^5$ <sup>43</sup>; and the effect of all these reagents in causing a profound diminution of double refraction of flow in myosin solutions (see p. 536).

The contrast between the effect of urea on serum albumin (Table 3) and on myosin (Table 4) is significant. The relatively compact serum albumin molecule apparently somehow uncoils, under the influence of urea, into a more elongated form, whereas myosin is broken up into smaller and less asymmetrical molecules. In both cases, it is probable that certain

TABLE 4. EFFECT OF DENATURING AGENTS ON VISCOSITY OF RABBIT MYOSIN AT HIGH VELOCITY GRADIENT

Myosin concentration,  $4.96 \times 10^{-3}$  gm. per cc. of solution =  $3.72 \times 10^{-3}$  cc. per cc. of solution

Substance added to 0.5 M KCl in solvent	$\eta_{sp}$	$\nu_{90}$
None.....	1.80	484
0.10 M glycine.....	1.80	484
1.5 M $\text{NH}_4\text{Cl}$ .....	1.45	390
0.62 M guanidine HCl.....	1.40	376
1.06 M guanidine HCl.....	1.04	279
2.14 M guanidine HCl.....	0.54	145
3.24 M guanidine HCl.....	0.36	97
2.5 M urea.....	0.84	226
5.0 M urea.....	0.34	91

From Edsall, J. T., and Mehl, J. W., *J. Biol. Chem.*, 133, 409 (1940).

relatively weak linkages within the molecule are ruptured under the influence of the urea. The studies of Anson and Mirsky<sup>44</sup> have indicated that the viscosity of solution of "globular" proteins, such as egg albumin, hemoglobin and edestin, is increased by reagents such as urea; whereas the measurements of Frampton and Saum<sup>45</sup> on tobacco mosaic virus indicate that urea has an effect on this substance qualitatively similar to its effect on myosin. (Compare the discussion in Chapter 19, p. 435.)

#### DOUBLE REFRACTION OF FLOW IN PROTEIN SOLUTIONS

It was discovered by Maxwell<sup>46</sup> in 1870 that Canada balsam, although isotropic at rest, becomes birefringent on flowing under shearing stress,

<sup>43</sup> Weber, H. H. and Stöver, R., *Biochem. Z.*, 259, 269 (1933).

<sup>44</sup> Anson, M. L. and Mirsky, A. E., *J. Gen. Physiol.*, 15, 341 (1932); see also Astbury, W. T., Dickinson, S. and Bailey, K., *Biochem. J.*, 29, 2351 (1935).



the birefringence disappearing as soon as the stress is removed. Similar observations on liquids were made about the same time by Ernst Mach in Vienna, and further studies of the phenomenon were made by other workers<sup>47</sup> in the following twenty-five years. The phenomenon was found to be particularly marked in certain types of colloidal solutions, and especially in sols of benzopurpurin and vanadium pentoxide, which were studied in great detail by Zocher, Freundlich and their collaborators<sup>48</sup>. The  $V_2O_5$  sols were shown to consist of needle shaped particles, which were small in freshly prepared solution, but on "aging" became greatly elongated. Correspondingly the amount of double refraction of flow increased greatly with time in any given preparation.

The first protein found to show well-marked double refraction of flow was myosin, the principal structural component of the muscle fiber<sup>49</sup>

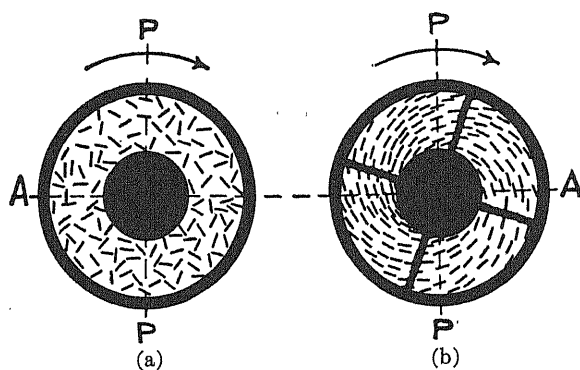


FIG. 4. Orientation of particles in a doubly refractive fluid placed between concentric cylinders; (a) particles, each schematically represented by a line indicating its optic axis, at rest; (b) orientation due to motion of concentric cylinders. From v. Muralt, A., and Edsall, J. T., *J. Biol. Chem.*, **89**, 315 (1930).

already discussed in this chapter because of its extremely high viscosity increment and the great change of the viscosity of its solutions with velocity gradient.

**Concentric Cylinder Apparatus; the Cross of Isocline and the Extinction Angle.** The form of apparatus employed by von Muralt and Edsall<sup>49</sup> was the same in principle as that used by many previous workers. Essentially it is a concentric cylinder apparatus, similar to the Couette viscometer, except that the inner cylinder is not suspended from a torsion wire, but is held rigidly fixed. Nicol prisms (or Polaroid discs) are placed above and below the annular space between the cylinders, in which the fluid moves,

and a parallel beam of light is passed through the system. If the Nicol prisms are crossed, and the liquid is at rest, the whole annular space remains dark (Fig. 4), since the molecules are oriented at random. If the outer cylinder is now set into motion, the very elongated myosin molecules are now partially oriented by the velocity gradients in the liquid. The annular space becomes bright, except for a dark cross—the “cross of isocline.” A crystal behaving in this fashion is called a spherocrystal; its optic axis is not linear, as with ordinary crystals, but is arranged with circular symmetry about a center. The stream lines in the liquid at any point are normal to the radius vector drawn from the center of the inner cylinder to that point. (The radius vector and the stream lines in Fig. 4 both, of course, lie in the plane of the paper). The optic axis of the element of liquid situated at that point does not in general coincide with the stream lines, but is inclined to them at an angle  $\chi$ , the extinction angle. Therefore at a point in the liquid which is separated from the plane PP or AA in Fig. 4 by the angle  $\chi$ , the optic axis is parallel to one of these planes. The light from the polarizer passes through the liquid at this point without change of its state of polarization, and is therefore extinguished by the analyzer. In other portions of the field, however, this is not true; the light is elliptically polarized, and part of it passes through the analyzer, so that the field remains bright<sup>50</sup>. Such a system may be characterized by the angle  $\chi$  between the stream lines and the optic axis of the flowing liquid; and by the magnitude of the double refraction,  $n_e - n_o$ . Here  $n_e$  is the refractive index for light polarized so that its electric vector vibrates parallel to the optic axis;  $n_o$  is the corresponding refractive index for light vibrating perpendicular to this axis. Values of these functions for rabbit myosin solutions of varying concentrations are shown in Fig. 5 and 6. In Fig. 5 the orientation of the optic axis is expressed, not by the angle  $\chi$ , but by its complement, the “angle of isocline,”  $\psi = 90^\circ - \chi$ . It is seen that in very dilute solutions and at low velocity gradients the optic axis approaches a position at  $45^\circ$  to the stream lines; at higher velocity gradients the optic axis shifts toward the stream lines, although it always in these experiments remained at least  $12^\circ$  from them. Boehm and Signer<sup>51</sup>, however, found that preparations of ovoglobulin show double refraction of flow, with the optic axis always parallel to the stream lines ( $\chi = 0^\circ$ ;  $\psi = 90^\circ$ ) even at low velocity gradients<sup>52</sup> and they concluded that the protein molecules responsible for the effect were very elongated. At the opposite extreme is the behavior of small molecules (alcohols, fatty acids, etc.) which show slight double refraction of flow at

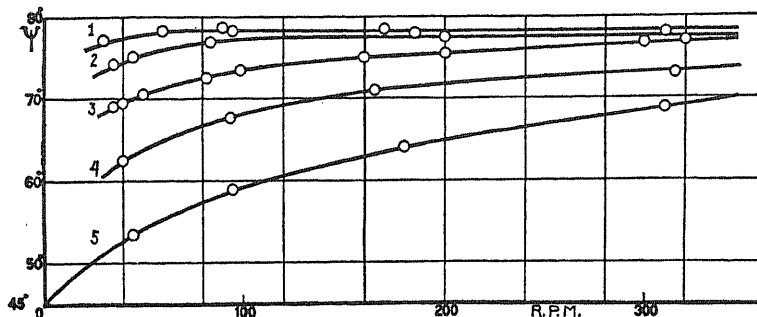


Fig. 5. The angle of isocline  $\Psi$  is plotted against the speed of rotation of the external cylinder for various concentrations of a preparation from a perfused rabbit. Curve 1, concentration 2240 mg. of N per liter,  $\eta'_{rel.}$  16.5; Curve 2, concentration 1790 mg. of N per liter,  $\eta'_{rel.}$  10.5; Curve 3, concentration 1345 mg. of N per liter,  $\eta'_{rel.}$  6.5; Curve 4, concentration 450 mg. of N per liter,  $\eta'_{rel.}$  2.2; Curve 5, concentration 225 mg. of N per liter,  $\eta'_{rel.}$  1.48. From von Mural, A., and Edsall, J. T., *J. Biol. Chem.*, 89, 315 (1930).

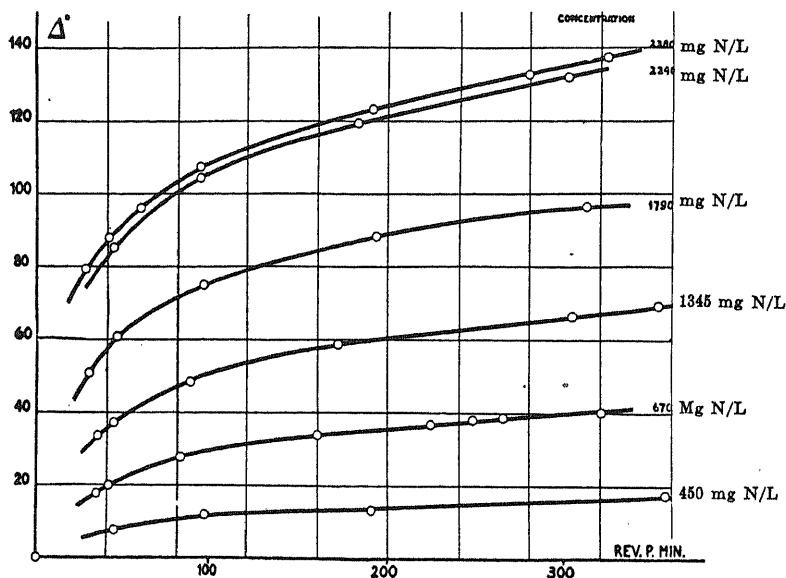


Fig. 6. Double refraction ( $\Delta$ ) of rabbit myosin as a function of velocity gradient, at several protein concentrations. From von Mural, A. L., and Edsall, J. T., *J. Biol. Chem.*, 89, 351 (1930).

very high velocity gradients<sup>53</sup>. In these systems invariably  $\chi = 45^\circ$  at all attainable velocity gradients.

The curves of double refraction as a function of velocity gradient (Fig. 6) rise steeply at first, and then become flatter, but never become horizontal within the region experimentally studied. On the other hand, the smaller substances for which  $\chi = 45^\circ$  give double refraction values which are directly proportional to the velocity gradient.

### Double Refraction of Ellipsoidal Molecules in Parallel Orientation

We may now consider the interpretation of results such as these, considering first the limiting case of a completely oriented system of parallel ellipsoidal molecules (refractive index  $n_1$ ) imbedded in a medium of refractive index  $n_2$ . Such a system was considered by Wiener<sup>54</sup>, who deduced that the system as a whole should be optically anisotropic even if both components are in themselves isotropic. For the case in which the imbedded molecules (or micelles) are small compared with the wave length of light, Wiener obtains, if the particles are elongated ellipsoids with long axes parallel, for the difference between the two refractive indices  $n_e$  and  $n_o$ :

$$n_e^2 - n_o^2 = \frac{\delta_1 \delta_2 (n_1^2 - n_2^2)^2}{(\delta_1 + 1)n_2^2 + \delta_2 n_1^2} \quad (40)$$

while if the ellipsoids are flattened platelets, with short axes parallel, the double refraction is negative:

$$n_e^2 - n_o^2 = - \frac{\delta_1 \delta_2 (n_1^2 - n_2^2)^2}{\delta_1 n_1^2 + \delta_2 n_2^2} \quad (41)$$

In these formulas  $\delta_1$  is the volume fraction of the parallel molecules,  $\delta_2$  that of the surrounding medium. Thus if  $n_1 = n_2$ , the system is isotropic, unless the molecules themselves are optically anisotropic. Systems of the type postulated by Wiener are found in many natural fibers, such as muscle, tendon, nerve and many plant fibers<sup>55</sup>; artificial protein fibers of the same sort are also readily produced by various techniques. Fig. 7 shows the double refraction of a gelatin fiber and of a myosin fiber, prepared by Weber<sup>56</sup>, when imbedded in media of different refractive indices,  $n_2$ . In both cases the double refraction of the fiber is a minimum near  $n_2 = 1.57$ , which may be taken as the refractive index of the protein. The gelatin fiber shows no double refraction at this minimum, indicating that the internal structure of the fiber is isotropic. The myosin fiber, however, is birefringent ( $n_e - n_o \cong 6 \times 10^{-3}$ ) even at the minimum, indicating an organized, anisotropic structure within the fiber molecules

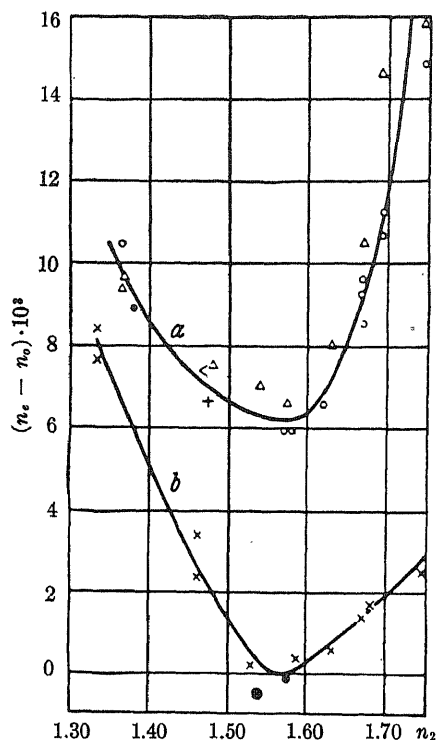


Figure 7.

Double refraction of myosin fiber (a) and gelatin fiber (b) as a function of the refractive index ( $n_2$ ) of the immersion medium. From Weber, H. H., *Arch. ges. Physiol.*, **235**, 205 (1934-35).

#### Theory of Double Refraction of Flow; Determination of Rotary Diffusion Constant

In solution, however, such anisotropic molecules are not all in parallel alignment, on account of their Brownian movement. In the presence of a velocity gradient, however, their distribution is not entirely random. The problem of calculating the angle  $\chi$  and the double refraction as a function of the velocity gradient and the rotary diffusion constant of the molecules was first successfully attacked by Boeder<sup>2</sup>. To simplify the problem, he treated in detail only the two dimensional case, in which the major axes of all the molecules lie in a single plane; also he confined the treatment to rod-shaped molecules of negligible cross dimensions (corresponding to the limiting case of an ellipsoid with the  $b$  axis = 0). It is necessary to evaluate the function  $\rho(\phi)$  defined in equation 1; that is, to calculate the relative numbers of molecules oriented at any given angle  $\phi$  to the stream lines. The situation of any molecule corresponds to that shown

number of particles having orientations within these limits is  $\rho(\phi_0)d\phi$ . If we denote by  $\omega$  the angular velocity of the particles due solely to the streaming of the liquid—neglecting, for the moment, the Brownian movement—then the number of particles which may be said to stream into this region at  $\phi_0$  in unit time, is  $(\rho\omega)_{\phi_0}$ . The number streaming out at  $\phi_0 + d\phi$  is  $(\rho\omega)_{\phi_0+d\phi}$ . The difference of these two expressions gives the time rate of alteration of  $\rho$  due to the streaming:

$$\left(\frac{\partial\rho}{\partial t}\right)_s = -\frac{\partial(\rho\omega)}{\partial\phi} \quad (42)$$

The angular velocity  $\omega$  is given by equation 34, taking  $b = 0$ , since the thickness of the molecule is negligible,

$$\omega = \frac{d\phi}{dt} = -\beta \sin^2 \phi \quad (43)$$

On the other hand, the change of  $\rho$  with time, due to the Brownian movement alone, is given by 3

$$\left(\frac{\partial\rho}{\partial t}\right)_b = \Theta \frac{\partial^2\rho}{\partial\phi^2} \quad (3)$$

The total rate of change of  $\rho$  with time at any point is given by the sum of 42 and 3

$$\frac{\partial\rho}{\partial t} = \Theta \frac{\partial^2\rho}{\partial\phi^2} - \frac{\partial(\rho\omega)}{\partial\phi} \quad (44)$$

If the system has reached a steady state, then we may set  $\frac{\partial\rho}{\partial t} = 0$  in 44.

In this case, the right hand side of (44) may be integrated to give

$$\Theta \frac{\partial\rho}{\partial\phi} - \rho\omega = \text{Const.} \quad (45)$$

or, dividing by  $\Theta$  and making use of 43:

$$\frac{\partial\rho}{\partial\phi} + \alpha\rho \sin^2 \phi = C \quad (46)$$

where we have, as previously in this chapter, set  $\alpha = \beta/\Theta$ . This differential equation was solved by Boeder, and the results are presented graphically in Fig. 8, in which  $\rho$  is plotted as a function of  $\phi$  for various values of  $\alpha$ . When  $\alpha$  is very small,  $\rho$  is found to be a maximum for  $\phi = 45^\circ$ , but the concentration of particles in this orientation is only slightly greater than for other values of  $\phi$ . As  $\alpha$  increases, the value of  $\phi$  at which  $\rho$  is a maximum

value of  $\rho$  becomes more and more pronounced; the increase in the observed value of double refraction is determined by the increasing height of this maximum. It is unnecessary to reproduce in detail the steps by which Boeder calculates the optical properties of the system from the orientation of the particles; the results are shown in Fig. 9, where  $\chi$  and the double refraction are shown as a function of  $\alpha$ . The maximum value of the double refraction, corresponding to complete orientation of the parti-

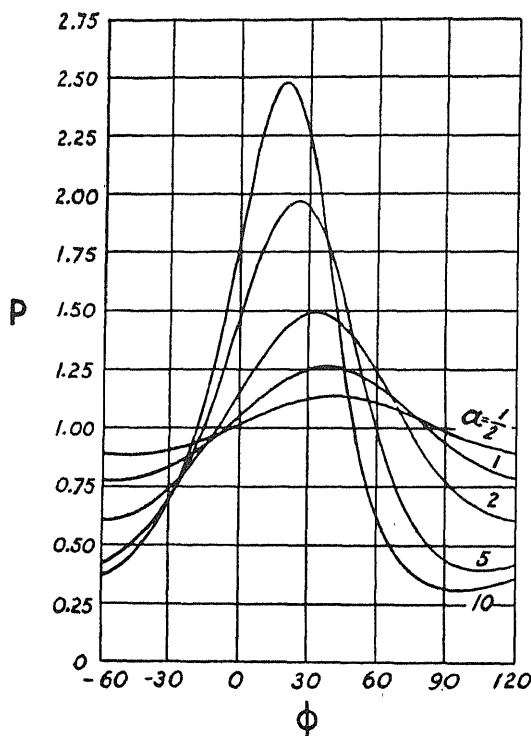


Fig. 8. The orientation distribution function,  $\rho$ , as a function of  $\phi$  for various values of  $\alpha = \beta/\theta$ . From Boeder, *Z. Physik*, **75**, 273 (1932).

cles, is arbitrarily set equal to  $2\pi^{57}$ . The values of  $\chi$  between  $\alpha = 0$  and  $\alpha = 15$  have been tabulated by Boeder (Table 5). Thus a single determination of  $\chi$  in a birefringent flowing liquid serves to determine a corresponding value of  $\alpha = \beta/\theta$ . Since the velocity gradient  $\beta$  is known from the dimensions of the apparatus and the angular velocity of the outer cylinder, this fixes the value of  $\theta$ . The length,  $2a$ , of the molecule may then be cal-

culated from equation 12b, for an elongated ellipsoid, taking a very roughly estimated value of  $a/b$ ; or the diameter  $2b$  for a flattened ellipsoid, using equation 16b.

Boeder's treatment is limited to thin, rod-shaped molecules, and his complete analysis is given only for the case of orientation in two dimensions. He has, however, given a less complete treatment for the problem of orientation in three dimensions. For low values of  $\alpha$  he derived the approximate formula for the extinction angle,  $\chi$  (expressed in radians):

$$\chi = \frac{1}{2} \tan^{-1} \frac{6}{\alpha} = \frac{\pi}{4} - \frac{\alpha}{12} \left( 1 - \frac{\alpha^2}{108} + \dots \right) \quad (47)$$

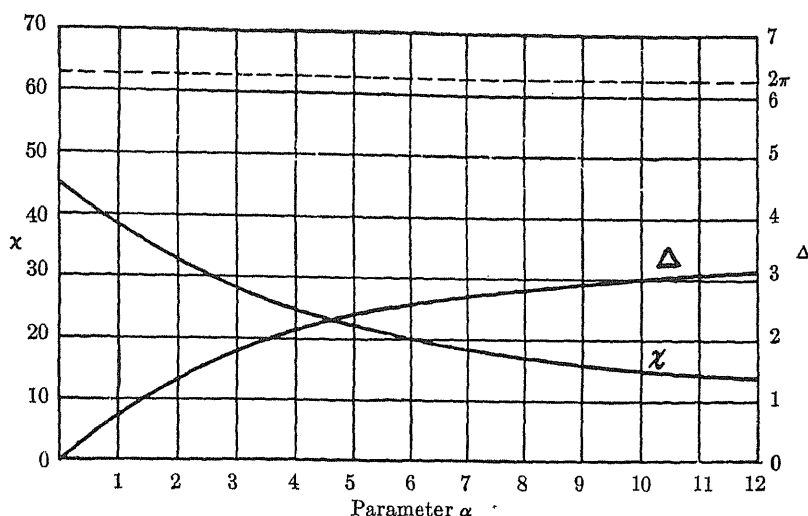


Figure 9. Extinction angle,  $\chi$ , and double refraction,  $\Delta$ , of a solution of thin rod-shaped particles. The limiting value of  $\Delta$ , corresponding to complete orientation, is set equal to  $2\pi$ . Abscissa:  $\alpha = \beta/\theta$ . Ordinate:  $\chi$  or  $\Delta$ . From Boeder, *Z. Physik.*, **75**, 273 (1932).

Recently Peterlin<sup>38</sup> and Peterlin and Stuart<sup>58</sup> have calculated the orientation of ellipsoids of revolution by a velocity gradient, and have developed a very complex mathematical expression for the orientation distribution function in three dimensions, as a function of  $\alpha$  and of the axial ratio ( $a/b$ ) of the ellipsoids. They obtain for  $\chi$  at low values of  $\alpha$ :

$$\chi = \frac{\pi}{4} - \frac{\alpha}{12} \left[ 1 - \frac{\alpha^2}{108} \left( 1 + \frac{24}{35} \frac{a^2 - b^2}{a^2 + b^2} \right) + \dots \right] \quad (48)$$



which for small  $\alpha$  values is seen to be practically identical with 47. When the extinction angle is not far from  $45^\circ$  these equations are preferable to the data of Table 5 as a basis for calculating the rotary diffusion constant from the extinction angle.

Since the calculated values of  $\Theta$  are very nearly inversely proportional to the cube of the molecular length, the error in the calculated length is much smaller than the error in  $\Theta$ <sup>59</sup>. We may employ equation 47 or 48 to calculate the length of the myosin molecule from the data shown in Fig. 5. For the most dilute solution studied,  $\psi \simeq 53^\circ$  and  $\chi \simeq 37^\circ$  at 47 rpm, corresponding to  $\beta = 10 \text{ sec}^{-1}$ . From Equation 47, the corresponding value of  $\alpha$  is very nearly 1.3. Hence  $\Theta \simeq 7 \text{ sec}^{-1}$ . Substituting this value in equation 12b, taking  $\eta$  as the viscosity of water, which is 0.016 at  $3^\circ$  (the temperature of the experiment), and setting  $a = 100b$  for calculation, we find  $a = 5800 \text{ \AA}$ ; the length of the molecule,  $2a$ , is thus about  $11600 \text{ \AA}$ , or very nearly the same as the length of the anisotropic band in a striated muscle fiber. Changing the estimated value of  $a/b$  by 500% alters the estimated length by only 10%<sup>60</sup>.

TABLE 5  
The Extinction Angle,  $\chi$ , as a Function of  $\alpha (= \beta/\Theta)$

$\alpha$	$\chi$	$\alpha$	$\chi$
0.0	$45^\circ$	3.5	$25^\circ 50'$
0.5	$41^\circ 27'$	5.0	$22^\circ$
1.0	$38^\circ 15'$	7.5	$17^\circ 30'$
2.0	$32^\circ 13'$	10.0	$15^\circ$

P. Boeder, *Z. Physik*, 75, 274 (1932).

### Effect of Chemical Agents on Double Refraction of Myosin

Double refraction of flow in myosin is reduced, to the point where it becomes undetectable at low velocity gradients, by many very mild chemical reagents. Von Muralt and Edsall<sup>49</sup> found that concentrated urea solutions, and more dilute solutions of iodides and thiocyanates, caused the double refraction to disappear; recently Edsall and Mehl<sup>41</sup> have shown that the same is true, not only of all the commonly recognized denaturing agents, but of many substances not commonly classed in this category. Among the latter, the chlorides of the divalent cations, calcium, magnesium and barium, were found to be particularly effective, causing disap-

<sup>59</sup> Werner Kuhn (*Z. physik. Chem. (A)*, 161, 427 (1932) and *Kolloid-Z.*, 62, 269 (1933)) has dealt with double refraction of flow arising from deformation of elastic molecules produced by the stresses in the flowing liquid. He gives a curve, Fig. 4 in the second paper cited, showing  $\chi$  as a function of  $\alpha$  in systems of this type. This curve is very different from that of Fig. 5, which is for rigid structures. The orientation of rigid curves. The  $\chi - \alpha$  curves so far obtained for proteins resemble Fig. 5 much more closely than they do Kuhn's curve. It is probable that this and other lines of evidence that proteins are relatively rigid structures, and that they play only a minor role in the phenomena discussed in this chapter. See also W. Haller,

pearance of double refraction within a few minutes at concentrations of 0.3 molar, or less. Ammonium, methylammonium and lithium salts had a similar action at somewhat higher concentrations. Guanidine hydrochloride, now recognized as a powerful denaturing agent for many proteins<sup>61</sup>, causes disappearance of double refraction at concentrations between 0.2 and 0.3 molar, while urea produces this effect only at concentrations several times as high. All these substances were effective at low temperatures and pH 7. In solutions alkaline to about pH 10.2 there was also a rapid disappearance of double refraction of flow; likewise in solutions acid to pH 5. These effects were found to be essentially alike for myosin from rabbit leg muscle and from lobster claw and tail muscle, and are therefore apparently characteristic of the myosins as a class.

These phenomena are best interpreted as reflecting a breakdown of the very elongated molecules of native myosin into smaller and less asymmetrical molecules—a conclusion borne out by the simultaneous fall in viscosity observed in such solutions (see Table 4). The solubility of the myosin in salt solutions is not diminished by the action of these reagents, and may even be increased. Although many of the same reagents produce striking effects upon the titratable sulfhydryl groups of myosin<sup>62</sup> there is no systematic correlation whatever between changes in double refraction of flow and changes in sulfhydryl groups. The nature of the labile linkages in native myosin which are split by these very mild reagents still remains to be discovered.

Very similar behavior has been found by Greenstein and Jenrette<sup>62a</sup>, who have studied the effect of various reagents on double refraction of flow in sodium thymonucleate. Guanidinium halides are particularly effective in this case also in producing a profound diminution of double refraction of flow. In sodium thymonucleate solutions, however, the double refraction is restored on removal of the agent which produced it, whereas the double refraction of myosin is not restored by washing out the denaturing agent<sup>62b</sup>.

### Tobacco Mosaic Virus

Notable among the proteins showing intense double refraction of flow is tobacco mosaic virus, which was first studied in this connection by Lauffer and Stanley<sup>63</sup>. This remarkable nucleoprotein, like myosin, is very sensitive to changes of pH and salt concentration; but unlike myosin,

<sup>61</sup> Greenstein, J. P., *J. Biol. Chem.*, **125**, 501 (1938); **128**, 233 (1939).

<sup>62</sup> Greenstein, J. P., and Edsall, J. T., *J. Biol. Chem.*, **133**, 397 (1940).

<sup>62a</sup> Greenstein, J. P., and Jenrette, W. V., *J. Nat'l Cancer Inst.*, **1**, 77 (1940); *Cold Spring Harbor Symp.*

it tends to pass into aggregation products which are larger and longer than the original untreated virus in the native state. Such changes are well shown by Figs. 10 and 11, taken from studies carried out by Mehl<sup>60</sup> on a virus preparation provided by Stanley. The original material, dissolved in phosphate buffer at pH 6.8, showed well marked double refraction (Curve 4 in Fig. 10), but double refraction increases very greatly when the pH is changed to 4.5 (Curve 1); and this change is only partially reversed when the pH is brought back to 6.8 (Curve 3). Fig. 11 shows the extinction angle in these same solutions as a function of the velocity gradient. For the original virus at pH 6.8, Mehl calculates a rotary diffusion constant

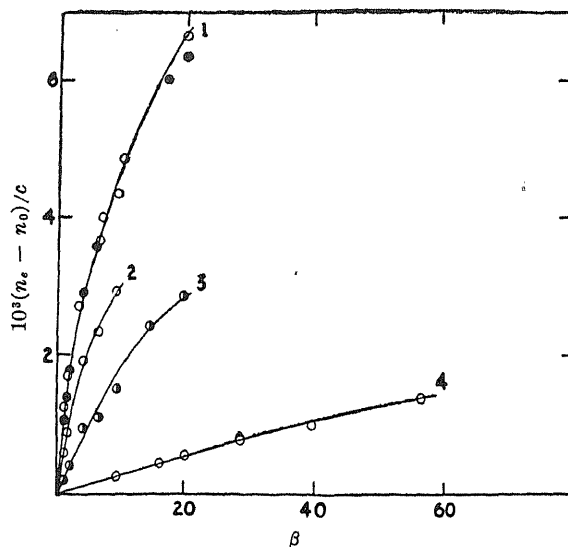


Figure 10. The double refraction of tobacco mosaic virus. Temperature 2–3°. Curve 1, open circles pH 4.5, closed circles pH 5.35. Curve 2, pH 6.0. Curve 3, pH 6.8, after having been at pH 4.5. Curve 4, original measurements at pH 6.8. From Mehl, J. W., *Cold Spring Harbor Symp. Quant. Biol.*, **6**, 218 (1938).

of  $25 \text{ sec}^{-1}$  at  $2^\circ$ , corresponding to a length of about  $5800 \text{ \AA}$ . In the solution at pH 4.5, the rotary diffusion constant is approximately 0.75, corresponding to a more than threefold increase in length. On returning to pH 6.8,  $\Theta$  rises to about  $3.4 \text{ sec}^{-1}$ , still far below the original value of  $25 \text{ sec}^{-1}$  at this pH. It is probable that even the molecules in the original solution at pH 6.8 were larger than molecules of the native virus, since this material had been prepared by chemical methods, while the later preparations obtained by the ultracentrifuge showed higher virus activity and apparently a lower state of aggregation<sup>64, 65, 66</sup>. It is fortunate that

direct estimates of the size and shape of this virus have now been obtained from studies with the electron microscope<sup>67</sup>. These revealed the existence of discrete rod-like particles, about 3300 Å in length and 150 Å in cross-section. The latter is in excellent agreement with the estimate of 150 Å for the cross-section made by Bernal and Fankuchen<sup>68</sup> from x-ray diffraction studies. (See Chapter 14). It is to be hoped that in future observations by all these different techniques, and others, may be made on a single sample of pure virus protein; since at present the observations of different workers on different preparations are often not strictly comparable. It is clear, however, that double refraction measurements can give valuable information about the lengths of these highly asymmetrical molecules, in harmony with results obtained by other methods.

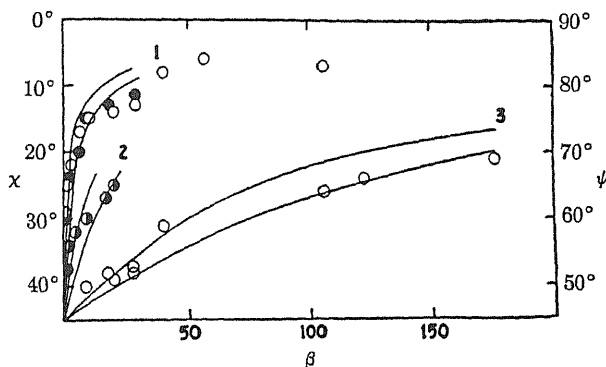


Figure 11. The extinction angle of tobacco mosaic virus as a function of the velocity gradient. Curve 1, open circles pH 4.5, closed circles pH 5.35, curves for  $\theta$  of 0.5 and of 1.0. Curve 2, pH 6.8 after having been at pH 4.5, curves for  $\theta$  of 3 and of 5. Curve 3, original measurements at pH 6.8, curves for  $\theta$  of 20 and of 30. From Mehl, J. W., *Cold Spring Harbor Symp. Quant. Biol.*, **6**, 218 (1938).

### Influence of Polydispersity

Boeder's treatment, on which our discussion has been based, assumes that the solute molecules are monodisperse. It has been shown by Sadron<sup>69</sup> that polydisperse systems may show very divergent behavior. In such systems, values of  $\beta$  may be found in which  $\chi$  increases, instead of decreasing, with increasing  $\beta$ . In a system composed of two constituents, with different values of  $\theta$  and opposite signs of birefringence for the particles,  $\chi$  may move across the stream lines, as the velocity gradient increases, and become less than 0°. Sadron has demonstrated experimentally the existence of such systems; for example a mixture of methylcellulose (posi-

<sup>67</sup> Krause, G. A., Fankuchen, E., and Burke, H., *Nature*, **177**, 800 (1955).

tive) and sodium thymonucleate (negative) reproduces semi-quantitatively the effects predicted by the theory.

TABLE 6. ROTARY DIFFUSION CONSTANTS AND APPROXIMATE LENGTHS OF CERTAIN PROTEINS AND OF SODIUM THYMONUCLEATE

Substance	Ref.	Temp. (°C)	$\Theta$ (sec <sup>-1</sup> )	Length, $2a$ in Angstroms (from Eq. 13)
Myosin (rabbit).....	(1)	3	7	11,600
Myosin (snail).....	(2)	25	ca. 1	ca. 28,000
Myosin (octopus).....	(2)	25	3, 5	18,000
Tobacco mosaic virus, pH 6.8..	(3)	3	25	7,200
Tobacco mosaic virus, pH 4.5..	(3)	3	0.75	24,000 ✓
Sodium caseinate (in 1.6N Na <sub>2</sub> SO <sub>4</sub> ).....	(4, 5)	20	700	2,200
Helix hemocyanin.....	(6)			890-960
Horse antibody globulin, M.W. 990,000.....	(6)			1,280
Fibrinogen.....	(7)	20	ca. 1200	ca. 1,800
Sodium thymonucleate.....	(8, 9) (13)	20	180	4,500

The lengths given here are not always those given in the original communications. Mehl (3, 2) and Nitschmann and Guggisberg (5) have used the equation of Werner Kuhn (10) for the length  $S$ :

$$\Theta = \frac{8KT}{\pi\eta S^3}$$

The results given here, however, are all still attended with some uncertainty. The value of  $a/b$  was taken as 100 for myosin, and for tobacco virus at pH 4.5, in applying equation 13; it was taken as 50 for tobacco virus at pH 6.8 and as 5.8 for Na caseinate in 1.6N Na<sub>2</sub>SO<sub>4</sub>, the latter being the value estimated (5) from viscosity measurements. The same value of  $a/b$  was arbitrarily chosen for fibrinogen and the value of 200 for sodium thymonucleate.

The sodium caseinate solution in 1.6N Na<sub>2</sub>SO<sub>4</sub> was certainly highly aggregated; many of the casein solutions studied by Nitschmann and Guggisberg (5) consisted of much smaller particles.

Tobacco mosaic virus has also been studied by Kausche, Guggisberg and Wissler (9) and by Robinson (11), with results in generally good agreement with those of Mehl (3).

For references to other proteins on which qualitative studies have been made, see Edsall (12).

The double refraction of all protein solutions yet studied is positive; that of sodium thymonucleate is negative.

(1) Murrill, A. v., and Edsall, J. T., *J. Biol. Chem.*, **89**, 315, 351 (1930); *Trans. Faraday Soc.*, **26**, 837 (1930).

(2) Mehl, J. W., *Biol. Bull.*, **79**, 483 (1940).

(3) Mehl, J. W., *Cold Spring Harbor Symp. Quant. Biol.*, **6**, 218 (1938).

(4) Nitschmann, H., *Helv. Chim. Acta*, **21**, 315 (1938).

(5) Nitschmann, H., *Helv. Chim. Acta*, **24**, 177 (1941).

(6) Sauerbrey, O., *Monatsh. Chem. Phys.*, **53**, 403 (1941).

(7) Sauerbrey, O., *Monatsh. Chem. Phys.*, **11**, 1 (1932).

(8) Sauerbrey, O., *Monatsh. Chem. Phys.*, **141**, 122 (1938).

(9) Kausche, H., *Monatsh. Chem. Phys.*, **117**, 303 (1939).

(10) Kuhn, W., *Z. Physik*, **2**, 61 (1933).

(11) Robinson, J., *Proc. Roy. Soc. London*, **A160**, 1 (1937). *Acids and Proteins*, C. C. Thomas, Springfield, Ill., 1939, p. 521.

(12) Wissler, A., Thesis, University of Bern (1940).

It has likewise been shown (70) that the

opposite signs of birefringence; the negative component is presumably of the nature of a lipid, since globulin treated with acetone and ether at 0° to remove lipoids behaves like a single component with positive birefringence. Globulin in aqueous salt solution, even without acetone-ether treatment, behaves like a single component, indicating that the lipid is bound to the globulin as an integral part of the molecule under these conditions. Treatment with glycerol presumably decomposes the complex.

### Studies on Other Proteins; the Lengths of Protein Molecules

Several serum albumin fractions were also studied by Sadron, Bonot and Mosimann.<sup>70</sup> Crystalline serum albumin, like egg albumin, was found to show no double refraction of flow even at the highest velocity gradients attained (around 40000 sec<sup>-1</sup>). A very soluble fraction, analogous to the glycoprotein of Hewitt, showed slight double refraction; and certain inter-

TABLE 7. LENGTHS OF CERTAIN PROTEIN MOLECULES FROM SEDIMENTATION AND DIFFUSION AND FROM DOUBLE REFRACTION OF FLOW

Protein	Molecular Weight	$f/f_0$	Length, 2a (Å) Sedimentation and Diffusion	Length, 2a (Å) Double Refraction
Hemocyanin ( <i>Helix</i> ).....	8,900,000	1.45	1130	890
Hemocyanin ( <i>Helix</i> ).....	4,300,000	1.40	820	890
Hemocyanin ( <i>Helix</i> ).....	1,030,000	1.79	820	960
Antipneumococcus serum globulin (horse).....	910,000	1.86	960	1280

Sedimentation and diffusion lengths from Sadron, Bonot, and Mosimann, *Compt. Rend. Acad. Sci. Paris*, 271, 1085 (1941).  
 Lengths from double refraction of flow, from Sadron, Bonot, and Mosimann, *Compt. Rend. Acad. Sci. Paris*, 271, 1085 (1941).  
 Lengths from double refraction of flow, from Sadron, Bonot, and Mosimann, *Compt. Rend. Acad. Sci. Paris*, 271, 1085 (1941).

mediate fractions showed still more; but all were much less birefringent than the globulin solutions under comparable conditions. These observations show that, with the very high velocity gradients available with Sadron's excellent apparatus, proteins which are only slightly asymmetrical may be studied by the method of double refraction of flow.

Determinations of double refraction of flow have now been made on several proteins with sufficient accuracy to allow an approximate calculation of length. It must be remembered in interpreting such data that nearly all the preparations hitherto studied are polydisperse. In consequence of this the values of the rotary diffusion constants determined at low viscosity gradients reveal mostly the influence of the longest particles present in the preparation. At higher viscosity gradients the influence of the shorter particles is more pronounced.

sion constants and lengths for a number of proteins are given in Table 6. The lengths as reported here are undoubtedly greater in almost all cases than the mean lengths of the molecules in the preparation, but they should give a very good idea of the approximate dimensions of the molecules involved. The values for hemocyanin and for horse antibody globulin are probably the most reliable as the preparations studied were also examined in the ultracentrifuge and found to be monodisperse. For these proteins it is possible to compare the lengths calculated from sedimentation and diffusion data with those calculated from double refraction of flow. The results are shown in Table 7. The study of *Helix* hemocyanin molecules of three different molecular weights was possible because of the reversible dissociation which this protein undergoes at different pH values and salt concentrations (Chapter 19, p. 439). It is interesting to note that the lengths of all three sizes of hemocyanin molecules are nearly the same; suggesting that the splitting of the largest molecule, when it dissociates, takes place parallel to the long axis. On the whole, the agreement between the lengths estimated by the two different methods is satisfactory, in view of the uncertainties involved in the measurements and calculations, from the  $f/f_0$  values and from double refraction of flow. Further studies on double refraction of flow in conjunction with studies of sedimentation, diffusion, dielectric dispersion and viscosity should yield increasingly accurate estimates of the shape of the proteins.

## Chapter 22

# The Electric Moments and the Relaxation Times of Proteins as Measured from their Influence upon the Dielectric Constants of Solutions

BY J. L. ONCLEY

Measurements of the dielectric constants of solutions of a fairly large number of proteins have been attempted, and give results which are of considerable importance. In Chapter 6 where the dielectric constants of amino acids and other simple dipolar ions were considered, it was pointed out that the dielectric constant is a function of the frequency of the electric field, and that measurements of dipolar moments require the use of sufficiently low frequencies to give dielectric increments which are independent of the frequency, and identical with the value determined in a static field. A brief discussion of the change of the dielectric constant with frequency called dielectric dispersion will be of aid in understanding certain problems which arise in the study of protein solutions.

### SIMPLE THEORY<sup>1</sup>

The dielectric constant,  $D$ , of any polar liquid or solution can be interpreted as being almost entirely a measure of the number of molecules oriented by an external electrical field of unit strength. These molecules are oriented by a torque depending on the field strength and the dipole moment,  $\mu$ , a constant for each molecular species. Orientation is hindered by the frictional forces in the solution depending on the frequency  $\nu$ , and a constant  $\tau$ , designated as the "relaxation time"<sup>2</sup>. Thus we find that the number of molecules oriented at unit field strength will decrease in the frequency region where the hindering frictional forces and the orienting torque become of the same order of magnitude. At lower frequencies the orienting torque is sufficient to overcome completely the resisting forces, and we have a high dielectric constant,  $D_0$ <sup>3</sup>. At very high frequen-



cies the resisting forces completely overcome the orienting torque and we again have a constant but low value of  $D$ .

If we now consider a binary mixture of two polar molecules (dipoles), we find a more complicated behavior. Fig. 1 represents the typical variation of dielectric constant with frequency for such a mixture when the relaxation times of the two components are quite different. The figure is divided into five regions, each of which exhibits a different dielectric behavior. In region A the orienting torque acting on dipoles of both types is sufficient to overcome all frictional forces and we find both being oriented

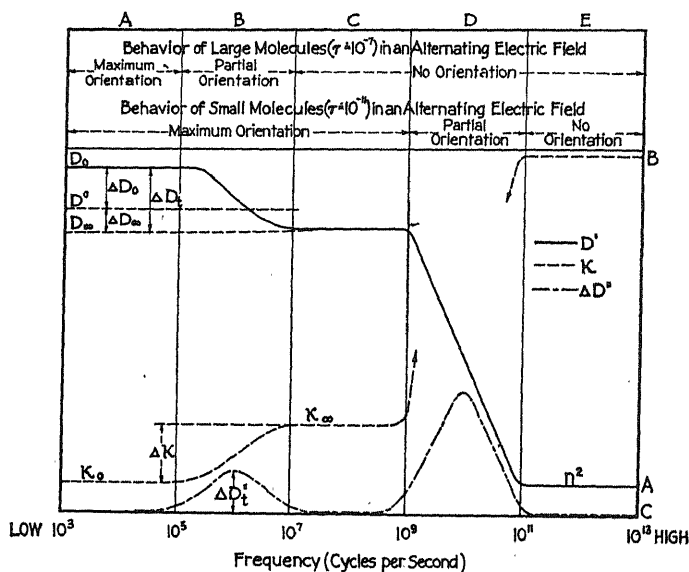


Figure 1. Schematic diagram of anomalous dispersion of the dielectric constant, ( $D'$ ), the specific conductivity ( $K$ ), and the dielectric absorption ( $\Delta D'' = 1.80(\kappa - \kappa_0)/\nu$ ) for two widely separated critical frequencies. From Shack, J., Ph.D. dissertation, Harvard University, 15, (1939). See also Oncley, J. L., Ferry, J. D., and Shack, J., *Ann. New York Acad. Sci.*, **40**, 371 (1940).

and a constant and high value of  $D$ , designated  $D_0$ . In region B, however, the frictional forces on dipoles with the larger relaxation time can no longer be neglected, the orientation of the larger molecules is no longer independent of the frequencies, and we find a region of decreasing dielectric constant. In region C this frictional force has overcome completely the orienting force in the case of the large molecules and they then contribute very little to the dielectric constant. The dipoles of smaller relaxation time, however, are still orienting independently of the frequencies, and we

D represents a region of decreasing orientation of these smaller molecules, and E a region of no orientation at all of molecules of either type. The dielectric constant here is very low, differing from unity only because of some energy stored in the dielectric by the distortion of the electron and nuclear positions in the molecules (atomic and electronic polarization), and this contribution is small when compared with that of the orientation of highly polar molecules.

In the case of aqueous solutions of the amino acids and peptides the regions B, C, and D are very close together, since the relaxation time for water is of the order of  $10^{-11}$  sec, while that of the amino acids and simpler peptides is  $10^{-10}$  sec. Further, it is very difficult to make measurements in any but region A, since a relaxation time of  $10^{-10}$  sec would not show much dispersion at frequencies below  $10^8$  cycles per second (wavelengths above 3 meters).

TABLE 1. DIELECTRIC INCREMENTS OF CERTAIN PROTEINS, AMINO ACIDS, AND PEPTIDES AT 25°C\*

Substance	Increment ( $\Delta D_0/g$ )	Substance	Increment ( $\Delta D_0/g$ )
$\alpha$ -Alanine.....	0.26	Egg albumin.....	0.1
Glycine.....	0.30	Glutadin.....	0.1
$\beta$ -Alanine.....	0.39	Serum albumin (carbohydrate-free).....	0.17
$\gamma$ -Aminobutyric acid.....	0.50	Insulin.....	0.3
Diglycine.....	0.54	Carboxyhemoglobin.....	0.33
$\epsilon$ -Aminocaproic acid.....	0.59	Zein.....	0.4
$\zeta$ -Aminoheptonic acid.....	0.59	Edestin.....	0.7
Triglycine.....	0.60	Saccharin.....	0.9
Heptaglycine.....	0.70	.....	1.1
Lysylglutamic acid.....	1.26	Lactoglobulin.....	1.5

\* Taken largely from Oncley, J. L., J. Phys. Chem., **44**, 1103 (1940).

When we consider molecules with much larger relaxation times, such as the proteins, we have quite a different situation. The proteins which have been studied have given relaxation times whose order of magnitude vary from  $10^{-6}$  to  $10^{-8}$  sec, so that regions B and D are quite clearly separated, and measurements can very frequently be made to cover regions A, B and C by using frequencies of from  $10^3$  or  $10^4$  to  $10^7$  cycles per second (wavelengths from 300,000 or 30,000 to 30 meters).

The low-frequency dielectric increments of a number of proteins, (as measured in region A) designated by the symbol  $\Delta D_0/g$ , have been recorded in Table 1. Here  $\Delta D_0 = D_0 - D^0$ , where  $D^0$  is the dielectric constant of the solvent, and  $g$  is the concentration of solute in grams per liter. Values of  $\Delta D_0/g$  for a few selected amino acids and peptides, as calculated from the equation  $\Delta D_0/g = \delta/M$  are also included in Table 1. The increments of

would, of course, be of a much larger order of magnitude for the proteins because of their high molecular weights.

### Calculation of Dipole Moments

The interpretation of these dielectric increments obtained in polar solvents has been discussed in Chapter 6. Most of the equations derived there could be equally well applied here, but since we are also in a position to obtain measurements of the "high frequency" increment per gram,  $\Delta D_{\infty}/g$ , in region C we can eliminate some of the approximations previously introduced, which is especially important for some proteins with small  $\Delta D_0/g$  values. We may write the equation

$$\mu^2 = [9000kT/4\pi Nh][M(\Delta D_0/g - \Delta D_{\infty}/g)] \quad (1)$$

where  $\Delta D_0/g$  and  $\Delta D_{\infty}/g$  are the dielectric increments per gram at very low (region A) and very high (region C) frequencies,  $N$  the Avogadro's number,  $k$  the Boltzmann's constant,  $T$  the absolute temperature, and  $\mu$  the electric moment of a single molecule in solution. This moment is not necessarily equal to the moment of a single molecule in the gas phase, but is in general somewhat greater<sup>4</sup>.

The parameter  $h$  may be evaluated in several different ways:

(1) Direct application of the Debye theory to polar media<sup>5</sup> yields a value  $h = (D_0 + 2)(D_{\infty} + 2)/3$ , which may be approximated by the equation  $h = (D^0 + 2)^2/3$  for solutions with sufficiently small increments.

(2) Studies by Wyman and others indicate that  $h = (D - j)/p$  in pure liquids and  $h = (D - D^0 - j)/p$  in binary mixtures, where  $p$  is the volume polarization and  $j$  another parameter. Wyman<sup>6</sup> suggests the values  $h = 8.5$  and  $j = -1$  as best representing the bulk of the available data.

(3) We may evaluate  $h$  by assuming values of  $(\Delta D_0/g - \Delta D_{\infty}/g)$  and  $\mu$  for the glycine molecule. Various measurements<sup>7</sup> indicate that  $\mu$  is close to 15, and we will choose this value. The total increment per gram may be taken as 0.38<sup>8</sup>.

(4) Onsager<sup>9</sup> has computed  $h$  as a function of an internal refractive index,<sup>10</sup>  $n$ , giving the result  $h = (n^2 + 2)^2/2 = 4.5[1 + (n^2 - 1)/3]^2$ .

<sup>4</sup> Kirkwood, J. G., *J. Chem. Phys.*, **7**, 911 (1939). See also Chapter 12.

<sup>5</sup> Oncley, J. L., *J. Am. Chem. Soc.*, **60**, 1115 (1938).

<sup>6</sup> Wyman, J., Jr., *J. Am. Chem. Soc.*, **58**, 1482 (1936); *Chem. Rev.*, **19**, 213 (1936).

<sup>7</sup> Kirkwood [*J. Chem. Phys.*, **2**, 351 (1934)] estimates the moment, 15.0 Debye units, on the basis of studies of the solvent action of neutral salts on glycine at low dielectric constants by Cohn [*Naturwissenschaften*, **20**, 663 (1932)]; and Scatchard and Prentiss [*J. Am. Chem. Soc.*, **56**, 2314 (1934)] estimate the moment, 14.8 Debye units, on the basis of freezing point measurements. A value near 15 is obtained for positive and negative charges in the glycine molecule (p. 15).

<sup>8</sup> J., Jr., and McMeekin, T. L. [*J. Am. Chem. Soc.*, **55**, 908

(1933)] where  $\bar{v}$  is the partial specific volume of glycine

(5) Kirkwood<sup>4</sup> has presented a method for computing  $h$  in terms of the hindered relative rotation of neighboring molecules. He gives the equation

$$h = 4.5 \left[ 1 + \frac{N}{v} \int_0^{v_0} \cos \gamma e^{-W/kT} d\omega dv \right] = 4.5(1 + \beta)$$

where  $\int_0^{v_0} e^{-W/kT} d\omega dv = 1$ . Here  $\gamma$  is the angle between the dipole moments of an arbitrary pair of molecules, and  $W$  is the potential between an arbitrary pair of molecules. The integration extends over all relative orientations ( $d\omega$ ) and positions ( $dv$ ) within a sphere of volume  $v_0$  which includes the molecule and the first few shells of neighboring molecules.

For the calculation of dipole moments it is convenient to modify equation 1 to become

$$\mu = \alpha \sqrt{M(\Delta D_0/g - \Delta D_\infty/g)} = \alpha \sqrt{M(\Delta D_i/g)} \quad (2)$$

TABLE 2. VARIOUS EVALUATIONS OF  $h$  AND  $\alpha$  FOR EQUATION 2 AND THE CORRESPONDING VALUES FOR THE DIPOLE MOMENT OF GLYCINE (7)

	$h$	$0^\circ$	$\alpha \times 10^{18}$ 20°	25°	Dipole Moment of glycine $\mu$	Onsager's refractive index $n$	Kirkwood's parameter $\beta$
Debye.....	2160	0.136	0.146	0.149	0.77	8.1	479.
Wyman.....	8.5	2.28	2.36	2.38	12.3	1.457	0.89
Empirical.....	5.8	2.76	2.86	2.89	15.0	1.182	0.29
Onsager or Kirk- wood <sup>a</sup> .....	4.5	3.13	3.24	3.28	17.0	1.000	0.0

<sup>a</sup> Minimum values for  $h$ .

Here a new parameter  $\alpha = \sqrt{9000kT/(4\pi N\hbar)} = 0.403 \times 10^{-18} \sqrt{T/h}$  has been introduced, and values of both  $h$  and this quantity at  $0^\circ$ ,  $20^\circ$ , and  $25^\circ$  are tabulated in Table 2. Values of  $n$  calculated from Onsager's equation and of  $\beta$  calculated from Kirkwood's equation are also given for the various values of  $h$ .

#### METHODS OF MEASUREMENT

The measurement of the dielectric constants of protein solutions over an extended frequency range can be carried out by any of several methods which have thus far been devised and employed, classifiable as (1) bridge methods, (2) resonance methods, (3) force methods, (4) calorimetric methods<sup>11</sup>, and (5) various other methods, involving comparisons of phase and magnitude of the voltage across an unknown and a standard.

(1) The bridge method involves the comparison of the resistance and capacitance of a cell containing the solution under investigation with some arrangement of standard resistances and capacitances. The main source

of difficulty is in the selection of a suitable resistance standard. Apparatus used for the study of protein solutions is described by Errera<sup>12</sup>, by Daniels, Mathews and Williams<sup>13</sup>, by Oncley<sup>4, 14</sup>, and by Ferry and Oncley<sup>15</sup>. Apparatus suitable for such measurements has also been described by Cole and Curtis<sup>16</sup>, and by Hemingway and McClendon<sup>17</sup>.

(2) The resonance method is widely used, especially for the higher range of frequencies (above  $10^6$  cycles per second). It involves the measurement of the resonance frequency of a circuit which includes a cell containing the solution. It has been widely used by Drude<sup>18</sup> and others. Many modifications have been proposed, the most accurate of which is probably that of Wyman<sup>19</sup>. Elliott and Williams<sup>20</sup> have recently described a method applied to protein solutions.

(3) The force method involves the measurement of the deflecting force exerted by an applied electric field upon a conducting ellipsoid suspended in the solution under investigation. It was originally described by Fürth<sup>21</sup> and has most recently been applied by Shutt<sup>22</sup> and his co-workers. It is most effective at the low frequencies ( $10^2$  to  $10^4$  cycles per second).

(4) The calorimetric method involves the measurement of the expansion of the solution as a measure of the heating produced by an applied field of high frequency. It has been used for the study of various solutions by Malsch<sup>23</sup>, Debye<sup>24</sup>, Martin<sup>25</sup> and others. Its application to protein solutions has been described by Shack<sup>26</sup>. It is most effective at the higher frequencies, but is capable of use over most of the usual dispersion range for protein solutions.

(5) Various other methods have been used for the measurement of the conductance and capacitance, or in some cases of the relaxation time directly, of a solution. Most of these methods depend upon the direct comparison of magnitude and phase of the voltage across an unknown and a standard capacitance, often by means of an oscillograph measurement. A recent publication of Marcy and Wyman<sup>26a</sup> describes such an apparatus.

- <sup>12</sup> Errera, J., *J. chim. phys.*, **29**, 577 (1932).
- <sup>13</sup> Daniels, F., Mathews, J. H., and Williams, J. W., "Experimental Physical Chemistry", McGraw Hill, New York, 3rd edition (1941), p. 378.
- <sup>14</sup> Oncley, J. L., *J. Phys. Chem.*, **44**, 1103 (1940).
- <sup>15</sup> Ferry, J. D., and Oncley, J. L., *J. Am. Chem. Soc.*, **63**, 272 (1941).
- <sup>16</sup> Cole, K. S., and Curtis, H. J., *Rev. Sci. Instr.*, **8**, 333 (1937); *Cold Spring Harbor Symp. Quant. Biol.*, **4**, 73 (1938).
- <sup>17</sup> Hemingway, A., and McClendon, J. F., *Physics*, **2**, 396 (1932).
- <sup>18</sup> Drude, P., *Z. physik. Chem.*, **23**, 267 (1897); **40**, 635 (1902).
- <sup>19</sup> Wyman, J., Jr., *Phys. Rev.*, **35**, 623 (1930).
- <sup>20</sup> Elliott, M. A., and Williams, J. W., *J. Am. Chem. Soc.*, **61**, 713 (1939).
- <sup>21</sup> Fürth, R., *Z. P.*, **12**, 111 (1929); *Handb. biol. Arbeitsmeth.*, IIIB, 835 (1929).
- <sup>22</sup> Shutt, W. J., *J. Phys. Chem.*, **30**, 893 (1934).
- <sup>23</sup> Shutt, W. J., and Hogan, H., *Proc. Roy. Soc. (London)* **A 157**, 359 (1936).
- <sup>24</sup> Dunning, W. J., and Shutt, W. J., *Trans. Faraday Soc.*, **34**, 479 (1938).
- <sup>25</sup> Carr, W., and Shutt, W. J., *Trans. Faraday Soc.*, **35**, 579 (1939).
- <sup>26</sup> Mason, W. A., and Shutt, W. J., *Proc. Roy. Soc. (London)* **A 175**, 234 (1940).
- <sup>23</sup> Malsch, J., *Physik. Z.*, **33**, 19 (1932); *Ann. Phys.*, **12**, 865 (1932).
- <sup>24</sup> Debye, P., *Trans. Faraday Soc.*, **30**, 679 (1934).
- <sup>25</sup> Martin, C., *Physik. Z.*, **33**, 19 (1932).
- <sup>26a</sup> Marcy, J., and Wyman, J., *J. Am. Chem. Soc.*, **64**, 1713 (1942).

Methods (3) and (4) just discussed (force and calorimetric) are more time-consuming methods than (1) and (2), and are most useful in the study of certain special cases. For routine measurements, on the other hand, methods (1) and (2) can often be so consolidated that a wide frequency range can be covered with the same cells and standards, or at least with a minimum of such equipment.

### Electrode Polarization

One of the most serious difficulties involved in measurements by any of these methods is the complication introduced by electrode polarization. This effect is most serious at low frequencies and high conductivities, and decreases in magnitude until it may properly be neglected at high frequencies or for solutions of very low conductivity. It is discussed in some detail by Cole and Curtis<sup>16</sup>, by Fricke and Curtis<sup>27</sup>, by Oncley<sup>4</sup>, by Ferry and Oncley<sup>15</sup>, and by Shutt<sup>22</sup> and must be considered as a possible source of error in all work of this type. Correction for this effect was made very successfully at frequencies lower than 25,000 cycles for salt solutions and protein solutions<sup>15</sup> by plotting capacities against  $\nu^{-3/2}$  and determining  $B$  empirically as the slope of the resulting straight line. The dielectric increment  $\Delta D$  is thus given by

$$\Delta D = (C_x - B\nu^{-3/2} - C_x^0)/Q \quad (3)$$

where  $C_x$  is the cell capacity at frequency  $\nu$  with the protein or salt solution,  $C_x^0$  the cell capacity with conductivity water, and  $Q$  the cell constant ( $dC_x/dD$ ). Below 25,000 cycles, however, the plot against  $\nu^{-3/2}$  often was found to deviate from linearity, and it was no longer possible to correct for the polarization capacity effect by this simple procedure. To involve the least additional complication, a method of comparison between solutions of equal conductivities has been employed. If the polarization correction is represented by  $B\nu^{-3/2} + f(\nu)$ , where  $f(\nu)$  takes care of the departure from linearity of the  $\nu^{-3/2}$  plot, and if  $f(\nu)$  should prove to be dependent only on the conductivity, but not on the nature of the solute, then the difference between the corrections of two solutions of different substances or mixtures with the same conductivity should be proportional to  $\nu^{-3/2}$ . The two constants  $B$  and  $B'$  will in general not be the same, but their difference,  $\Delta B$ , can be determined from the plot of  $C_x - C_x'$  against  $\nu^{-3/2}$ , and the values of  $\Delta D$  obtained as

$$\Delta D = (C_x - C_x' - \Delta B\nu^{-3/2})/Q \quad (3a)$$

This method has been tested by a series of measurements comparing solutions of potassium chloride, ammonium sulfate, glycine (with potassium chloride added to adjust the conductivity) and bovine serum albumin.

## DIELECTRIC INCREMENTS

## Low-frequency Dielectric Increments

The low-frequency dielectric increments of most proteins behave in a manner similar to those of the amino acids and peptides (Chapter 6). A

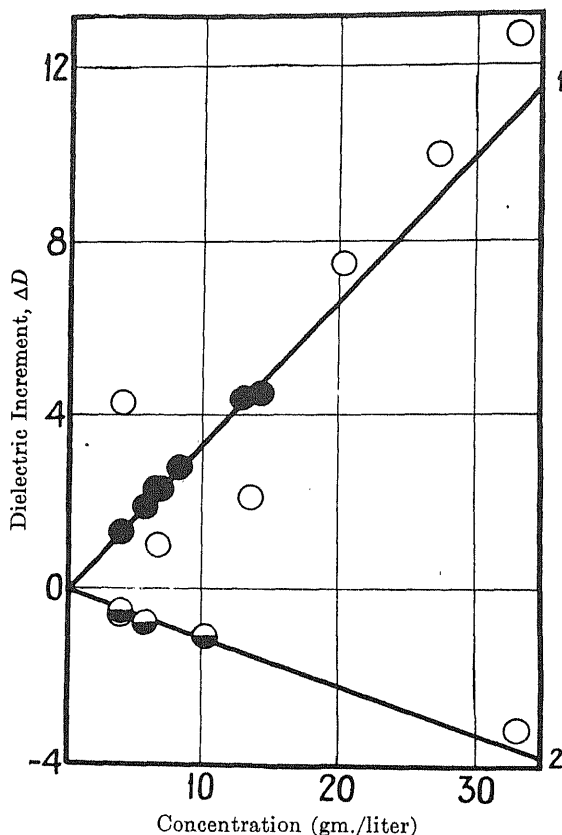


Figure 2. Low (1) and high (2) frequency dielectric increments of carboxyhemoglobin:

○, Errera, *J. Chim. Phys.*, **29**, 577 (1932); ●, Wyman, J., Jr., found  $\Delta D_{\infty}/g$  values of  $-0.11$ ,  $-0.14$ , and  $-0.11$  for solutions of concentrations 4.0, 5.8, and 10.2 grams per liter, and the value  $\Delta D_{\infty}/g = -0.11$  is used in this calculation. ●, Oncley, *J. Am. Chem. Soc.*, **60**, 1121 (1938). From Oncley, J. L., *J. Am. Chem. Soc.*, **60**, 1115 (1938).

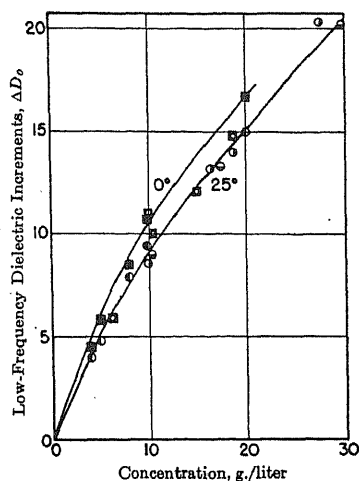
linear relationship between the increment  $\Delta D_0$  and the concentration, usually expressed in grams per liter,  $g$ , has been observed in dilute solution, so that  $\Delta D_0/g$  at low concentrations is a constant, about  $0.11$  for

$\Delta D_0/g$  is large, the increment per gram decreases somewhat (Fig. 3). Thus we may write

$$\Delta D_0 = E_0 g - F_0 g^2 \quad (4)$$

where  $E_0$  and  $F_0$  represent parameters characteristic of the individual proteins. Terms in higher powers of  $g$  may sometime be required, but present indications are that the coefficients for these terms will be small, and negligible except in very concentrated solutions. The observed decreases in the dielectric increment per gram,  $\Delta D_0/g$ , in the case of very polar molecules, suggests that this effect may be due to electrostatic interactions between these molecules of high dipole moment. Ferry and Oncley<sup>15</sup> found that the parameter  $F$  in equation 4 was in qualitative agreement with the calculations of Fuoss<sup>28</sup> for the decrease of polarization through electro-

Figure 3.  
Dielectric increments of horse serum pseudo-globulin- $\gamma$ .



static interaction; the decrease should be greater for molecules of greater dipole moment and at lower temperatures.

The effect of temperature upon the parameters  $E_0$  and  $F_0$  of equation 4 has been less thoroughly investigated. The coefficient  $\alpha$  of equation 2 is seen to be proportional to the square root of the absolute temperature, so that a constant dipole moment would require that  $T(\Delta D_i/g)$  be constant. This is not quite true for the cases which have been studied (lactoglobulin and  $\gamma$ -pseudoglobulin), as can be seen by a reference to Table 3. These differences, however, are small, and may be within experimental error. On the other hand, these differences might be real, and due to a change in



The change of low-frequency increment with pH has been studied for only a few protein molecules. Shutt<sup>22a</sup> has published results obtained when the pH of egg albumin solutions was varied from about 4.1 to 5.9 by the addition of sodium hydroxide or sulfuric acid. The increments he obtained are illustrated in Fig. 4, showing minima at pH 4.4 and 5.2, and a maximum in the isoelectric region. This central portion of the curve is similar to those obtained for various amino acids<sup>22c, 22d</sup>. Studies on horse serum  $\gamma$ -pseudoglobulin have been made by Oncley, Bateman, Pecher and Melin<sup>29</sup> with the pH varying between 5.0 and 7.5 due to the addition of hydrochloric acid or potassium hydroxide (Fig. 5). No maxima or minima were observed in this case, and the increment varied almost linearly with the pH, being largest in alkaline solutions. We cannot calculate dipole mo-

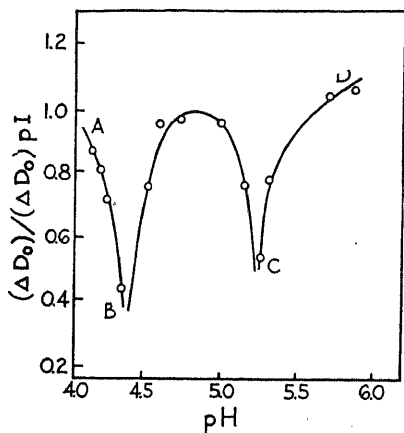


Figure 4.

Variation of low frequency dielectric increment ( $\Delta D_0$ ) of egg albumin with pH. The value of  $\Delta D_0$  at the isoelectric point (pI) is set equal to unity in this figure. 2.5% egg albumin, temperature 20°C.

Data from Shutt.<sup>22a</sup>

ments for these solutions at pH values other than isoelectric, however, since it is necessary to know the "center of symmetry" of a charged ion before its dipole moment can be evaluated. A charged ion undergoes not only a translational movement when placed in an electric field, but also a rotational movement if the center of symmetry of the ion does not coincide with the center of charge.<sup>29a</sup> This effect gives rise to an increased electric moment, probably best described as the moment we would calculate for the electrically neutral assembly of charges obtained by considering the ion (with a charge of  $n$  electronic units, say) and an additional charge of  $-n$  electronic units located at the center about which the ion rotates (its "center of symmetry"). Dipole moments calculated in the usual manner for such charged particles would accordingly be too large by a

factor whose magnitude depends upon the total charge of the ion, the location of the charge, and the symmetry of the ion. These new factors are for the most part unknown, so that we must be satisfied at present to calculate "apparent dipole moments". In the case of the horse serum  $\gamma$ -pseudoglobulin at 0° these vary from 1100 Debye units at pH 5 to 1400 at pH 7.5.

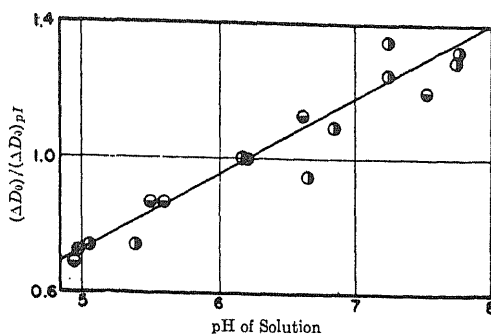


Figure 5. Ratio of low-frequency dielectric increments of 1% horse serum  $\gamma$  pseudoglobulin solutions of varying pH to the isoelectric value.

### High-frequency Dielectric Increment

The high-frequency dielectric increment has been measured for only a few of the proteins which have been studied. These values are recorded in Table 3. If we assume that the dielectric constant at high frequencies,  $\Delta D_\infty$ , is due largely to the contribution of solvent molecules (assuming the volume occupied by the protein to have a high-frequency dielectric constant of unity<sup>30</sup>), then we obtain the equation

$$-\Delta D_\infty/g = (D^0 - 1)v/1000 \quad (5)$$

where  $v$  is the volume of water (in cc) displaced by one gram of anhydrous protein. We can set  $v = \bar{v} + w/\rho_0$ , where  $\rho_0$  is the density of the solvent,  $\bar{v}$  the partial specific volume of the solute (anhydrous), and  $w$  the number of grams of water which appear to be associated with each gram of anhydrous protein. If  $w = 0$  and  $\bar{v} = 0.75$ , we obtain a value of  $-D_\infty/g = 0.058$  for water at 25°, and 0.065 at 0°. The observed values are of this order of magnitude, but slightly larger. If we assume reasonable values for  $w$ <sup>31</sup>, the agreement is very good. In the cases where the low-frequency increments are quite large, values of the high-frequency increments are often estimated from equation 5 using  $v = 0.75$ .

<sup>30</sup> The value of unity for the high-frequency dielectric constant of the protein is too small since it neglects

There are as yet no data available as to the variation of the high-frequency increment with temperature, solvent, pH or concentration. The

TABLE 3. DIELECTRIC INCREMENTS AND DIPOLE MOMENTS OF VARIOUS PROTEIN MOLECULES

Protein	Solvent	Temp. (°C.)	$E_0$	$F_0$	$-E_\infty$	$E_t$	$M$	$\mu$
Horse carboxyhemoglobin....	Water	25	0.33	0	0.09	0.42	67,000	480
Pig carboxyhemoglobin.....	Water	20				0.3	(67,000)	410
Myoglobin..	Water	25	0.15		(0.06)	0.21	17,000	170
Insulin.....	80% aqueous propylene glycol	25				0.38	40,000	360
Insulin.....	90% aqueous propylene glycol	25				0.29	40,000	310
Insulin.....	100% propylene glycol	25				0.26	40,000	300
$\beta$ -Lactoglobulin.....	$M/2$ and $M/4$ glycine	25	1.51	0.025	(0.07)	1.58	40,000	730
$\beta$ -Lactoglobulin.....	$M/2$ and $M/4$ glycine	0	1.84	0.047	(0.08)	1.92	40,000	770
Egg albumin.....	Water	25	0.10	0	0.07	0.17	44,000	250
Horse serum albumin (carbohydrate-free)	Water	25	0.17	0	0.07	0.24	70,000	380
Horse serum $\gamma$ -pseudo-globulin...	Water	25	1.08	0.017	(0.06)	1.14	142,000	1100
Horse serum $\gamma$ -pseudo-globulin...	Water	0	1.26	0.023	(0.06)	1.32	142,000	1300
Edestin.....	2M glycine	25	0.7		(0.1)	0.8	310,000	1400
Gliadin.....	56% aqueous ethanol	25				0.10	42,000	190
Secalin.....	54% aqueous ethanol	25				1.0	24,000	440
Zein.....	72% aqueous ethanol	25				0.45	40,000	380

symbol  $E_\infty$  will sometimes be used to represent the value of  $\Delta D_\infty/g$  at zero concentration, just as  $E_0$  represented the similar value of  $\Delta D_0/g$ .

#### Total Dielectric Increment

In cases where proteins are dissolved in mixed solvents, there is sometimes difficulty in exactly defining the composition of the solvent since one

and  $\Delta D_0/g$ , and we must be content with measurements of the total increment,  $\Delta D_i/g$ .

#### DISPERSION OF DIELECTRIC CONSTANT AND CONDUCTANCE

##### Dielectric Dispersion

The results which have been discussed have been obtained by the study of regions A and C as defined in Fig. 1. The study of the region B in which dielectric dispersion occurs is capable of giving us information of an entirely different type. This dispersion has been found to be dependent upon the size and shape of the molecules, and gives us one of the most powerful methods for the accurate measurement of these quantities that is at present available.

The quantitative behavior of the dielectric constant in this dispersion region is somewhat complex. Debye<sup>32</sup> has treated the simplest case involving a single relaxation time, and obtained the equations

$$D' = D_\infty + (D_0 - D_\infty)/(1 + \nu^2/\nu_c^2) \quad (6)$$

Here  $D'$  is the usual dielectric constant of the solution (often called the "real" dielectric constant),  $D_0$  and  $D_\infty$  are the dielectric constants observed in regions A and C, respectively,  $\nu$  is the frequency, usually expressed in cycles per second, or megacycles per second, and  $\nu_c$  is the "critical frequency", defined as  $1/(2\pi\tau)$ , where  $\tau$  is the relaxation time, discussed in Chapter 21.

This simple theory, involving a single relaxation time and critical frequency, can be extended to include the case of several relaxation times, giving

$$D' = D_\infty + \Delta D_1/(1 + \nu^2/\nu_1^2) + \Delta D_2/(1 + \nu^2/\nu_2^2) + \dots \quad (7)$$

where  $\Delta D_1$ ,  $\Delta D_2$ , etc., represent the total dielectric increments associated with the critical frequencies  $\nu_1$ ,  $\nu_2$ , etc. The critical frequencies are defined as before; that is,  $\nu_i = 1/(2\pi\tau_i)$ . It was shown in Chapter 21 that an ellipsoidal molecule would in general have three relaxation times, but if we restrict ourselves to ellipsoids of revolution, then we have only two relaxation times,  $\tau_a$  and  $\tau_b$ , and hence two critical frequencies,  $\nu_a$  and  $\nu_b$ . Equations of Perrin (equations 12-16, Chapter 21) give  $\tau_a$  and  $\tau_b$  in terms of the ratio of axes of the ellipsoid,  $a/b$ , and the relaxation time  $\tau_0$  of a sphere of the same volume, and we can thus calculate  $\nu_a$  and  $\nu_b$  in terms of  $a/b$  and  $\nu_0$ . If  $\Delta D_i$  be taken as the sum of  $\Delta D_a$  and  $\Delta D_b$ , then equation 7 can be expressed in terms of  $a/b$ ,  $\nu_0$ ,  $\Delta D_i$ , and  $\Delta D_a/\Delta D_b$ , and we can construct series of curves of  $(D' - D_\infty)/\Delta D_i$  against  $\nu/\nu_0$  for various values of  $\Delta D_a/\Delta D_b$ .

are illustrated in Figs. 6 and 7, and numerical values are recorded in Tables 1 to 8 of the appendix (p. 645). By comparison of the experimental curves obtained in region B (Fig. 8) with theoretical curves of this type, we can

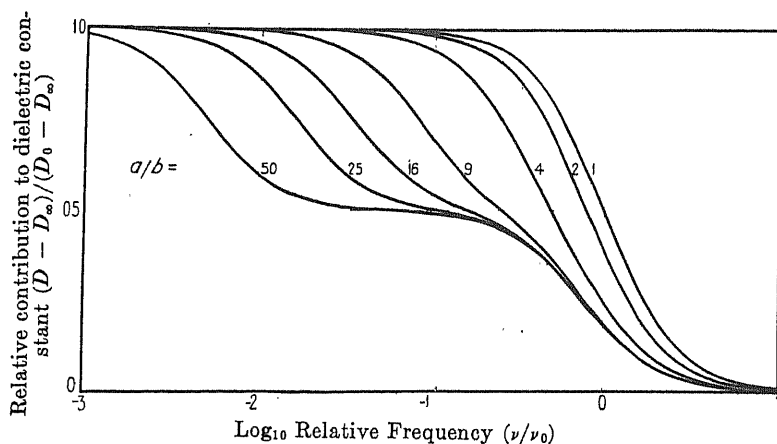


Figure 6. Dielectric dispersion curves for elongated ellipsoids of revolution (according to Perrin), with  $\Delta D_a/\Delta D_b = 1$  and  $\theta = 45^\circ$ .

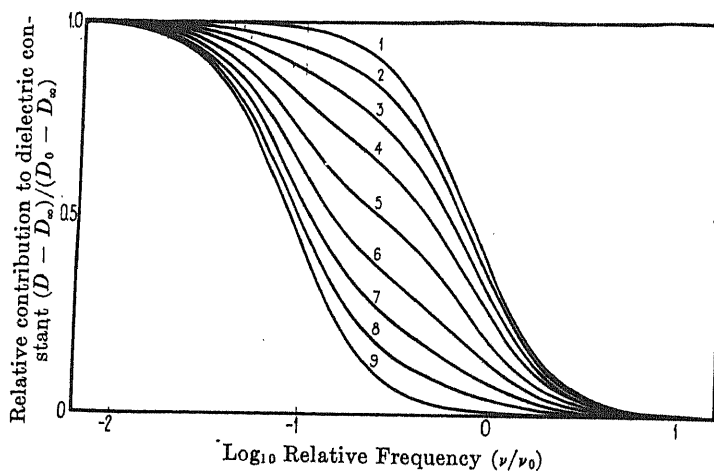


Figure 7. Dielectric dispersion curves for ellipsoids of revolution with  $a/b = 9$ . Curves 1-9 are for  $\Delta D_a/\Delta D_b = 0, 0.1, 0.25, 0.5, 1, 2, 4, 10$ , and  $\infty$ , respectively.

then evaluate  $a/b$ ,  $\Delta D_a/\Delta D_b$ , and  $\nu_0$ . The values of  $\nu_0$  (or  $\tau_0$ ) are directly related to the molecular volumes,  $V$ , of the particles:

$$\nu_0 = kT/(8\pi^2 ab^2 \eta) = RT/(6\pi V \eta) = 1/(2\pi \tau_0) \quad (8)$$

or

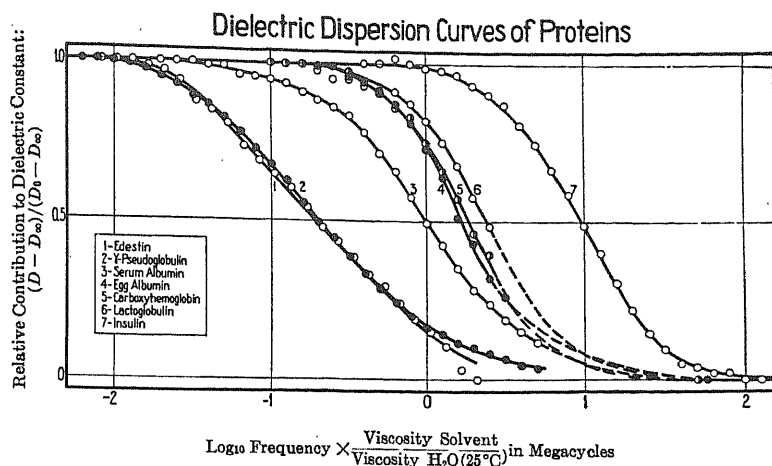


TABLE 4. CRITICAL FREQUENCIES, RELAXATION TIMES AND GEOMETRIC ASYMMETRY OF VARIOUS PROTEIN MOLECULES

Protein	Solvent	Temp. (°C.)	$\frac{\eta}{\tau_{25,w}}$	$\nu_{obs.}$	$\frac{\Delta D_1}{\Delta D_2}$	$\tau_{25,w} \times 10^8$	$\tau_{0,25,w} \times 10^8$	$\frac{a}{b}$	$\theta$ degrees
Horse carboxyhemoglobin.....	Water	25	1.00	1.9		8.4	(6.6)	(1.6)	
Pig carboxyhemoglobin.....	Water	20	1.12	1.1		13.0			
Myoglobin.....	Water	25	1.00	5.5		2.9			
Insulin.....	80% aqueous propylene glycol	25	17.0	0.59		1.6			
Insulin.....	90% aqueous propylene glycol	25	28.0	0.31		1.8			
Insulin.....	100% propylene glycol	25	48.0	0.19		1.8			
$\beta$ -Lactoglobulin.....	M/4 glycine	25	1.04	1.0; 3.0	0.25	15; 5.1	4.3	4	63
$\beta$ -Lactoglobulin. ...	M/2 glycine	25	1.08	0.9; 2.6	0.25	16; 5.7	4.8	4	63
$\beta$ -Lactoglobulin.....	M/2 glycine	0	2.12	0.44; 1.25	0.25	16; 5.5	4.6	4	63
Egg albumin.....	Water	25	1.00	0.86; 3.4	1.5	18; 4.7	3.7	5	40
Horse serum albumin (carbohydrate-free).....	Water	25	1.00	0.44; 2.1	1.0	36; 7.5	6.0	6	45
Horse serum pseudoglobulin- $\gamma$ .....	Water	25	1.00	0.064; 0.57	1.0	250; 28	22.0	9	45
Horse serum pseudoglobulin- $\gamma$ .....	Water	0							
Edestin.....	2M glycine	25	1.33	0.060; 0.44	1.0	240; 27	21.0	9	45
Gliadin.....	56% aqueous ethanol	25	2.58	0.23; 1.6	1.6	27; 3.8	3.1	8	38
Secalin.....	54% aqueous ethanol	25	2.59	0.21; 2.3	1.5	29; 2.7	2.1	10	40

For water at 25°, this becomes

$$V = 147,000/\nu_0^{25,w} = 9250\tau_0^{25,w} \times 10^8, \quad (9a)$$

where  $\nu_0$  is expressed in megacycles, and  $\tau_0$  in seconds. The increment ratio,  $\Delta D_a/\Delta D_b$ , may also be expressed as a "dipole angle",  $\theta$ , defined as the angle between the geometric axis  $a$  of the ellipsoid and the electric moment vector. Since the electric moment  $\mu$  will be the vector sum of

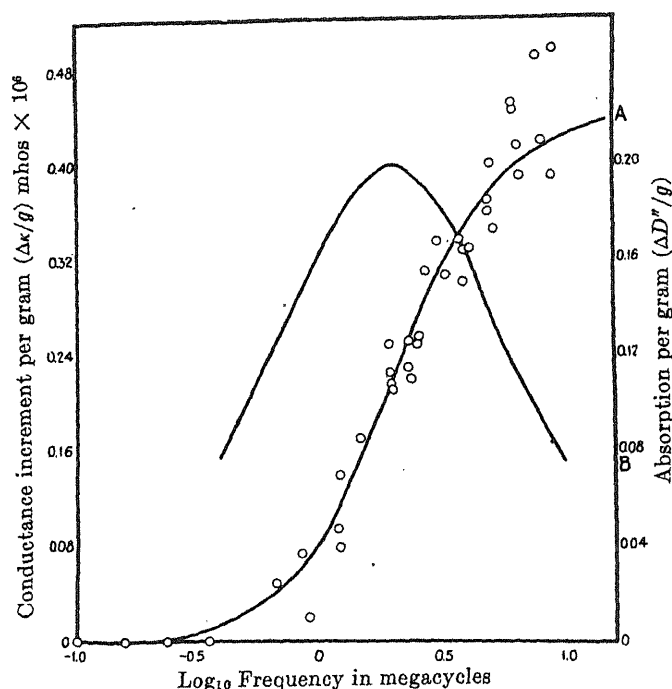


Figure 9. Dielectric absorption of carboxyhemoglobin solutions: C, experimental measurements of  $(\kappa - \kappa_0)/g$ ; curve A,  $(\kappa - \kappa_0)/g$ , and curve B,  $\Delta D''/g$ , calculated from equation 1 of Oncley, Ferry and Shack, using the constants, recorded in Table 1 of their paper, for the calorimetric method. [Shack, J., Ph.D. Dissertation, Harvard University 50 (1939)]. From Oncley, J. L., Ferry, J. D., and Shack, J., *Annals New York Acad. Sci.*, **40**, 371 (1940).

the two moments  $\mu_a$  and  $\mu_b$ , and since the moments are proportional to the square roots of the increments, we have

$$\tan \theta = \mu_b/\mu_a = \sqrt{\Delta D_b/\Delta D_a}. \quad (10)$$

#### Dispersion of Conductance

Measurements of the specific conductance,  $\kappa$ , of a protein solution can

TABLE 5. COMPARISON OF RESULTS OBTAINED ON SEVERAL PROTEIN MOLECULES STUDIED BY BOTH BRIDGE AND CALORIMETRIC METHODS

Protein	Method	Critical Frequency $\nu_c$	Dielectric Increment $\Delta D_i/g$	Conductance Increment $\Delta \kappa/g$
Egg albumin (8-9% conc.) 1st relaxation	Bridge	0.88	0.092	(0.045)
	Calorimetric	.80	(0.102)	0.046
	Best values	.86	0.094	0.045
Ditto, 2nd relaxation	Bridge	3.6	0.071	(0.142)
	Calorimetric	3.3	(0.062)	0.114
	Best values	3.4	0.063	0.119
Ditto, total increments	Bridge		0.163	(0.187)
	Calorimetric		(0.165)	0.154
	Best values		0.157	0.164
Carboxyhemoglobin (horse)	Bridge	1.9	0.44	(0.46)
	Calorimetric	2.0	(0.40)	0.44
	Best values	1.9	0.42	0.44
$\beta$ -Lactoglobulin (0.1% in water)	Bridge	2.4	1.5	(2.0)
	Calorimetric	2.2	(1.4)	1.7
	Best values	2.4	1.5	2.0

Taken in part from Oncley, Ferry, and Shack, *Annals New York Acad. Science*, 40, 383 (1940).

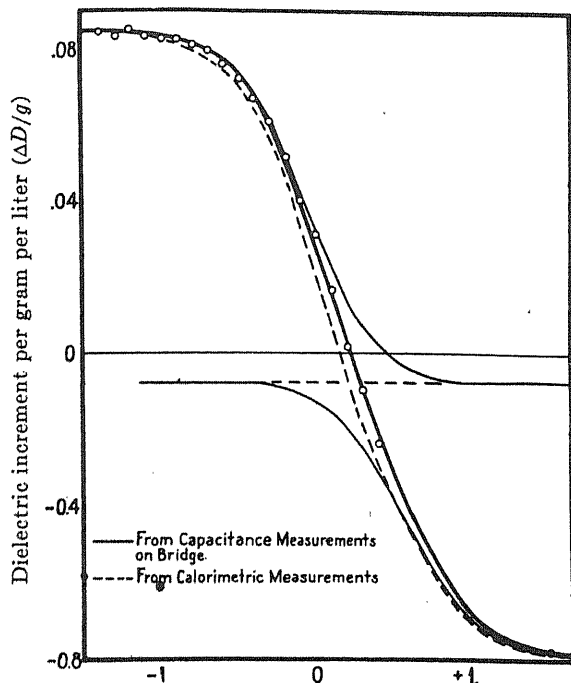


Figure 10.

Dielectric increments for egg albumin solutions. From Oncley, J. L., Ferry, J. D., and Shack, J., *Annals N. Y. Acad. Sci.*, 40, 371 (1940).



will have the value  $\kappa_0$ , which increases in region B to the value  $\kappa_\infty$  in region C. We thus have the equations

$$\kappa = \kappa_\infty - (\kappa_\infty - \kappa_0)/(1 + \nu^2/\nu_c^2) = \kappa_0 + (\kappa_\infty - \kappa_0)(\nu/\nu_c)^2/(1 + \nu^2/\nu_c^2) \quad (11)$$

and

$$\kappa = \kappa_\infty - \Delta\kappa_1/(1 + \nu^2/\nu_1^2) - \Delta\kappa_2/(1 + \nu^2/\nu_2^2) - \dots \quad (12a)$$

$$= \kappa_0 + \Delta\kappa_1(\nu/\nu_1)^2/(1 + \nu^2/\nu_1^2) + \Delta\kappa_2(\nu/\nu_2)^2/(1 + \nu^2/\nu_2^2) + \dots \quad (12b)$$

to correspond to equations 6 and 7 for the dielectric constant. Here  $\Delta\kappa_i$  and  $\nu_i$  represent the conductance increment and critical frequency of the  $i$ 'th dispersion region. The relation

$$\Delta\kappa_i = \Delta D_i \nu_i / 1.80 \quad (13)$$

exists between the conductance increment, the critical frequency, and the dielectric increment of any dispersion region when conductances are expressed in  $\mu\text{mhos/cm}$  and critical frequencies in megacycles. The results of a comparison of conductance increments, (Fig. 9) dielectric increments, and critical frequencies by independent bridge and calorimetric methods are recorded in Table 5. The mean values in this table are weighted, more weight being given to the low-frequency bridge data and to the high-frequency calorimetric data. Values enclosed in parentheses are calculated from equation 13. Figure 10 also compares the data obtained by these two methods in the case of egg albumin.

## RESULTS

A survey of the dielectric constant literature for protein solutions reveals few data of significance before those of Wyman<sup>33</sup>. Before this time most workers had made measurements at such high frequencies that the results obtained fell in region C (Fig. 1) where the contribution of the protein molecules to the dielectric effect was very small; and even these measurements were very uncertain. The development of new techniques of measurement since 1928 have made observations in regions A and B much more reliable. The results discussed in the following paragraphs are confined almost entirely to these more recent investigations.

### Carboxyhemoglobin (horse)

Measurements of Oncley<sup>5</sup> and of Shack<sup>26</sup>, made at temperatures near 25° indicate that the dispersion curves are characterized by a single critical frequency and high- and low-frequency increments per gram which are independent of the concentration (Figs. 2 and 8 and Tables 3, 4 and 5). These measurements were further shown to be completely reproducible

years, and were in fair agreement with earlier results obtained by Errera<sup>12</sup>. The magnitude of the electric moment calculated for this molecule (Table 3) shows that carboxyhemoglobin has a fairly large dielectric effect. The observed critical frequency corresponds to a relaxation time somewhat too large to be attributed to a spherical molecule with a molecular volume of 50,000 cc ( $66,700 \times 0.749$ ), but can be explained by assuming an elongated ellipsoid of axial ratio  $a/b$  of about 1.6 and molecular volume of about 60,000 cc (corresponding to a hydration of about 0.2 gram of water per gram of protein) (see Fig. 11).

#### Carboxyhemoglobin (pig)

Dielectric constant measurements on this molecule have been made by Arrhenius<sup>34</sup>. He found that the dielectric increments varied considerably with different preparations of hemoglobin, giving extremes of 0.21 and 0.10 for the total increment per gram at protein concentrations around 4%. He also finds some variation of total increment with concentration, values at low protein concentration (under 1%) giving increment values considerably larger than those obtained at high protein concentrations. Measurements were also made upon solutions containing glucose, glycerin, urea and glycine. The dispersion curves obtained for the pig hemoglobin could be explained by a single relaxation time, but the times so obtained were somewhat larger than those obtained for horse carboxyhemoglobin (Table 4).

#### Myoglobin

Myoglobin, prepared from horses' hearts, has been studied by Marcy and Wyman<sup>26a</sup> by the use of an oscillograph method which enables them to measure both the dielectric increment and the conductance increment at various frequencies. The electric moment for myoglobin is quite small, due both to its low increment and its low molecular weight. The observed critical frequency cannot be attributed to a spherical molecule with no hydration, but can be explained quite well by assuming hydrations of 0.2–0.3 gram of water per gram of protein, and slightly asymmetrical shape.

#### Insulin

Insulin, which is almost insoluble in water at the isoelectric point, has been studied in mixtures of water and propylene glycol<sup>35</sup>. The dielectric increment varies somewhat with composition of solvent, increasing with increasing water content. The dispersion curves, when analyzed for single

to the viscosities of the respective solvents (within 10%). They are shown in Fig. 8 and Table 4. The relaxation time so obtained is, however, too

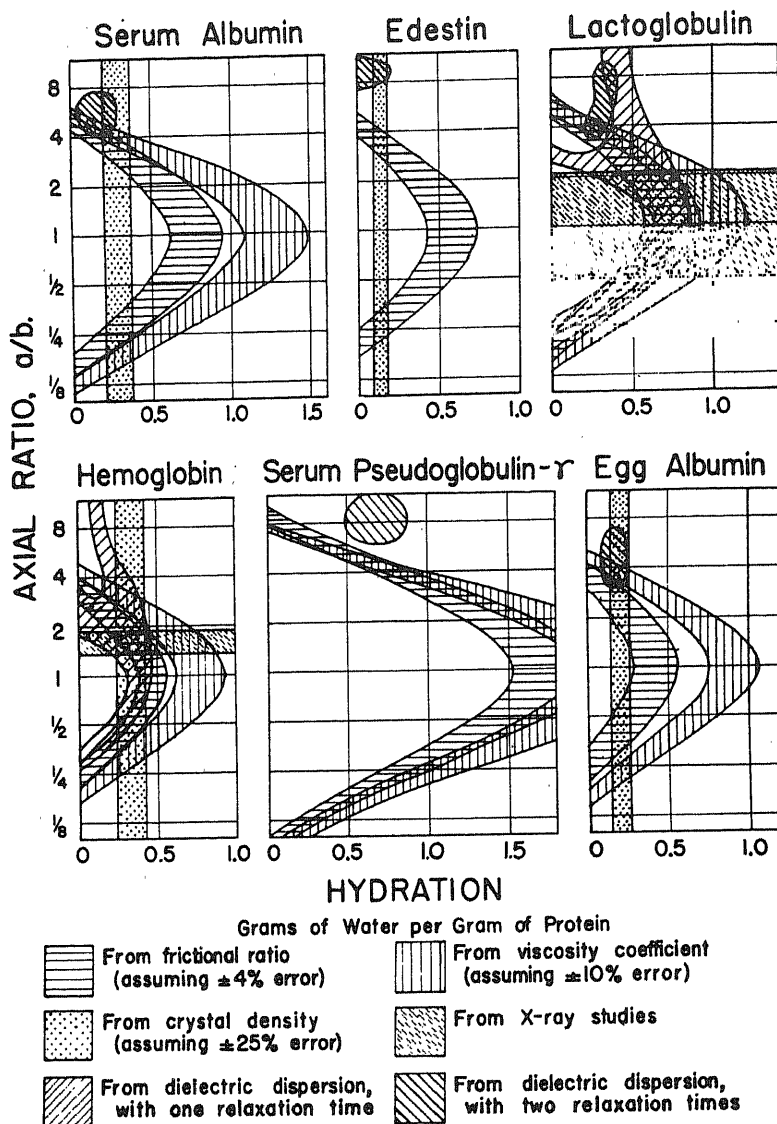


Figure 11. Asymmetry and hydration of certain protein molecules

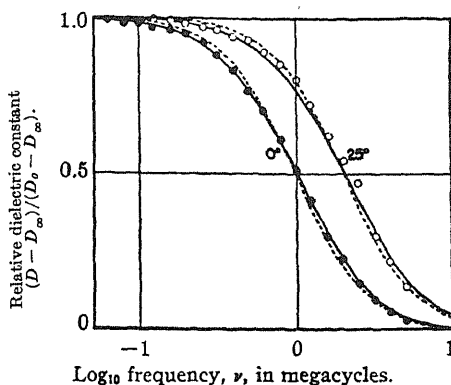
small to be interpreted as due to the rotation of a spherical or ellipsoidal

**$\beta$ -Lactoglobulin**

Preliminary measurements were reported by Ferry, Cohn, Oncley, and Blanchard<sup>36</sup>, and later results by Ferry and Oncley<sup>15</sup> and Shack<sup>26</sup>. The measurements were made in  $M/2$  and  $M/4$  aqueous glycine solutions at both 0 and 25°. The increment per gram obtained was the largest yet observed for any protein, although the electric moment obtained was somewhat lower than that of serum pseudoglobulin and edestin. The observed dispersion curves are plotted in Figure 12. The continuous curves are calculated on the basis of the two relaxation times of an elongated ellipsoidal molecule with  $a/b = 4$ ,  $\tau_0^{25,w} = 4.7 \times 10^{-8}$  sec, and  $\theta = 63^\circ$ . The broken curves are calculated from the simple Debye expression for a single relaxation time,  $\tau^{25,w} = 7.5 \times 10^{-8}$  sec. Both sets of curves fit the points reasonably well; however, the continuous curves fit somewhat better, espe-

Figure 12.

Dielectric dispersion of lactoglobulin dissolved in 0.5  $M$  glycine.  $\circ = 25^\circ$ ;  $\bullet = 0^\circ$ . (—) continuous lines, calculated from two relaxation times for an elongated ellipsoid,  $(\tau_0)H_2O$ ,  $25^\circ = 4.7 \times 10^{-8}$  sec.,  $a/b = 4$ ,  $\theta = 63^\circ$ ; broken lines, calculated from one relaxation time,  $5H_2O$ ,  $25^\circ = 7.5 \times 10^{-8}$  sec.



cially the one at  $0^\circ$ . Since the measurements at  $0^\circ$  are more reliable because of (1) the lower conductivity of the solutions at this temperature, and (2) the location of the dispersion at lower frequencies, the interpretation based on the two relaxation times of an elongated ellipsoidal molecule seems somewhat more probable. Results of Shack indicated a system characterized to the precision of the measurements by a single dispersion region. Table 5 compares the values for increments and frequencies as obtained from dielectric constant and specific conductance measurements.

**Egg Albumin**

Egg albumin solutions have been studied by Shutt<sup>22a</sup>, Oncley<sup>37</sup> and Shack<sup>26</sup>. Here the experimental determination of the dispersion data is difficult, however, since the dielectric constant increment is small. The results reported in Tables 3, 4 and 5 and Fig. 10 are for a fairly concen-

trated solution of about 9% egg albumin. More dilute solutions are being studied. The relaxation times so obtained can be attributed to an ellipsoidal molecule of axial ratio  $a/b = 4$  and molecular volume about 34,000 cc. If we take 42,000 for the molecular weight, this gives a hydration of about 0.1 gram of water per gram of protein.

#### Serum Albumin (horse)<sup>38</sup>

Measurements of dielectric increment have been made by Errera<sup>12</sup>, Ferry and Oncley<sup>39</sup>, and Oncley<sup>14</sup>. The measurements of Errera were of preliminary nature, and upon preparations of undetermined properties. The measurements of Ferry and Oncley were made upon several preparations whose solubility in ammonium sulfate varied by a factor of about five, and indicated that the total dielectric increments of the more soluble fractions were approximately three times greater than the least-soluble fractions, varying from 0.42 to 0.13 dielectric unit per gram per liter. All preparations gave very nearly identical dispersion curves which were not exactly reproduced by equation 6 with one relaxation time. It was concluded that "there remains the possibility that further purification may yield a product whose dispersion shows no deviation from the simple Debye theory". The most recent measurements indicate that this possibility does not seem very likely, since measurements upon the carbohydrate-free albumin of McMeekin<sup>40</sup> indicate two relaxation times (Table 4). The total increment per gram per liter obtained was about 0.24, (Table 3) and the relaxation times were interpreted as agreeing well with an axial ratio  $a/b$  of 5 or 6 and a hydration of about 0.2 gram of water per gram of protein (Fig. 11).

#### Serum Pseudoglobulin (horse)<sup>38</sup>

Measurements on serum pseudoglobulin were made by Ferry and Oncley<sup>39</sup>, who found that these solutions had a very high dielectric increment per gram which depended somewhat upon the concentration. The dispersion curve indicated two relaxation times. Measurements of Oncley<sup>14</sup> upon  $\gamma$ -pseudoglobulin confirmed these observations. He found two relaxation times corresponding to a somewhat greater asymmetry than observed for the total pseudoglobulin (Fig. 8). An axial ratio,  $a/b$ , of about 8 or 9 and a hydration of about 0.5 gram of water per gram of protein would explain the observed dispersion curves fairly accurately (Fig. 11). Fig. 3 shows the dielectric increments observed for these new preparations at both 0 and 25°.

<sup>38</sup> Measurements upon albumin and pseudoglobulin fractions obtained by ammonium sulfate precipitation of human urine from a case of ... (Oncley, D. G., Ferry, J. D., and Oncley J. L. [J. Biol. Chem., 123, Proc. xxxix ...] per gram obtained for the pseudoglobulin was the same as obtained for horse ... for the albumin was somewhat lower ...)

### Edestin

Measurements of the dielectric increments of edestin dissolved in 2*M* glycine solutions have been reported by Oncley<sup>12</sup> (Fig. 8). The observed total increment of about 0.8 unit per gram per liter indicates quite a polar molecule. The dispersion curve was almost identical with that observed for serum  $\gamma$ -pseudoglobulin, and revealed two relaxation times compatible with an elongated ellipsoidal model with axial ratio,  $a/b$ , of about 9 and but little hydration.

### Gliadin

This protein has been studied by Arrhenius<sup>41</sup>, who found that the total dielectric increment per gram was about 0.15. He reported only a single relaxation time, and stated that lower frequencies would need to be employed for the measurements to reveal a second relaxation time. Recent measurements of Entrikin<sup>42</sup> have included such low frequency results, and yield a total dielectric increment per gram of about 0.10, and two relaxation times compatible with an elongated ellipsoidal model with axial ratio,  $a/b$ , of about 8 and molecular weight of about 38,000.

### Secalin

Measurements were made by Andrews<sup>43</sup> on the protein secalin from rye. He obtained a total increment per gram of about 1.0, and two relaxation times which would indicate an axial ratio,  $a/b$ , of about 10 and a molecular weight of about 40,000.

### Zein

Preliminary measurements on solutions of zein in 70% *n*-propyl alcohol were made by Wyman<sup>33</sup>, who found a total increment per gram of about 0.3. Measurements of Elliott and Williams<sup>44</sup> have been carried out in 72% ethanol solutions, and over a considerable temperature and concentration range. They find a total increment per gram of about 0.46, and two relaxation times which give an axial ratio,  $a/b$ , of about 7. Some differences between "whole zein" and "fractionated zein" were observed.

### Amino Acids and Peptides<sup>45</sup>

The first definite indications of dispersion in amino acid solutions were obtained by Fricke and Parts<sup>46</sup>, who made measurements at frequencies up

<sup>41</sup> Arrhenius, S., *J. Chem. Phys.*, **5**, 63 (1937).

<sup>42</sup> Entrikin, P. P., *J. Am. Chem. Soc.*, **63**, 2127 (1941).

<sup>43</sup> Andrews, A. C., *J. Am. Chem. Soc.*, **62**, 942 (1940). See also Ph.D. Dissertation, Univ. of Wisconsin (1938).

<sup>44</sup> Elliott, M. A., and Williams, J. W., *J. Am. Chem. Soc.*, **61**, 718 (1939). See also Williams, J. W., and

to 66 megacycles. They obtained no change of dielectric increment with frequency in this range, but they did observe considerable change in conductance. From this, they estimated the critical frequency for glycine (at 21°) as  $3.5 \times 10^9$  cycles/sec, in a 1*M* solution, corresponding to a relaxation time of  $4.6 \times 10^{-11}$  sec. Larger amino acids gave longer relaxation times. Recently Bateman and Potapenko<sup>47</sup> have measured dielectric increments and absorption coefficients of amino acids at 1176 megacycles ( $\lambda = 25.5$  cm). The dielectric increments were lower than in a static field, the lowering being especially marked in the solutions of the larger molecules, while the dielectric absorption increases with size. Relaxation

TABLE 6. RELAXATION TIMES OF AMINO ACIDS AND PEPTIDES

Substance	Relaxation time, $\tau_{25}^{10} \times 10^{11}$				
	Fricke and Parts*	Bateman and Potapenko†	Wyman and Marcy‡	Conner and Smyth§	Calculated for rigid sphere¶
Glycine.....	3.7	2.6		4.9	6.2
$\alpha$ -Alanine.....	5.2	6.6		6.4	7.9
$\beta$ -Alanine.....		6.7			7.9
$\alpha$ -Aminobutyric acid.....	6.2				9.7
$\gamma$ -Aminobutyric acid.....	7.9				9.7
$\alpha$ -Aminocaproic acid.....	9.7				13.2
$\epsilon$ -Aminocaproic acid.....	12.9				13.2
Diglycine.....		14.9		11.4	10.1
Glycylalanine.....		20.3			11.8
Alanylglycine.....		21.3			11.8
Leucylglycine.....				18.0	17.2
Leucylalanine.....				24.6	18.8
Lysylglutamic acid.....			48.0		22.8
Triglycine.....			13.0	18.1	14.0
Alanyldiglycine.....				20.6	15.8
Leucyldiglycine.....				28.2	21.1
Alanylleucylglycine.....				27.8	22.8
Tetraglycine.....				24.7	17.0
Pentaglycine.....				36.3	21.9

\* Calculated from data of Fricke and Parts at 65.6, 32.8, and 16.4 megacycles, using equations 5, 11, 13.

† Calculated from measurements at 1180 megacycles. (Temp. 23.3°C.)

‡ Calculated from measurements at 115 megacycles.

§ Calculated from measurements at 375 to 750 megacycles.

¶ Calculated from equation 9, using apparent molal volumes and electrostrictions from Chapter 7.

times calculated from these measurements are recorded in Table 6. The value for glycine,  $2.6 \times 10^{-11}$  sec, is not far from the earlier estimate of Fricke and Parts. In the last column of Table 6 are listed the calculated relaxation times,  $\tau_0$  of spheres of the same molecular volume as the amino acids and peptides studied, as calculated from (8). The observed values are on the whole remarkably close to the calculated values. Still more recent results of Marcy and Wyman<sup>26a</sup> at 115 megacycles have been used

and indicated very little change in dielectric increment, but considerable change in conductance. Relaxation times at about 25° of  $4.8 \times 10^{-10}$  for lysylglutamic acid and  $1.3 \times 10^{-10}$  for triglycine were obtained by an application of equations 6 and 11. Values calculated from equation 8 for spheres of the proper molar volume for these molecules were 2.4 and  $1.4 \times 10^{-10}$ , respectively, and they attribute the larger observed value in the case of lysylglutamic acid to asymmetry of the molecule.

Since this manuscript was prepared, Conner and Smyth<sup>48</sup> have reported measurements at frequencies from 375 to 750 megacycles of the dispersion of a number of amino acids and peptides, and these results have been entered in Table 6. They agree with other workers in observing relaxation times for glycine and alanine which are less than the values calculated for a spherical model, and conclude that "this discrepancy is customary for investigations of this sort for small molecules of size comparable with those of the solvent, and has led to the suggestion that the inner viscosity is of smaller magnitude than the macroscopic viscosity". On the other hand, the peptides they studied all had relaxation times somewhat larger than the values calculated for spheres and these results were used to calculate the axial ratio,  $a/b$ , of the assumed molecular ellipsoid of revolution, as determined from Perrin's equations<sup>49</sup>. Values of  $a/b$  vary from 1.1 to 2.1 for these peptides. From these values for the glycine peptides, Conner and Smyth conclude that "although the accuracy of the determinations is not sufficient to exclude the possibility of a linear increase with the first power (of the chain length), the apparent linear dependence upon the square root of the number of glycine residues is what one would expect from a statistical consideration of rigid molecules randomly distributed in all possible configurations resulting from potential minima symmetrically distributed about the valence bonds of the back bone chain" (Ref. 48, p. 1878. See also ref. 50). They also discuss the values of observed relaxation times in terms of the free energy of activation for rotation of these molecules as obtained following Eyring's treatment<sup>51</sup>.

#### CONCLUSIONS

The studies of dielectric increments and dispersion of these various proteins have added considerably to our knowledge of the geometrical and electrical symmetry of protein molecules. The most important results which have been obtained may be summarized as follows:

Dipole moments of the proteins which have been studied (from 170 to 1400 Debye units) seem very large when compared with those of low molecular weight substances, but this is due largely to the high molecular



weights of the proteins. A comparison of the electrical symmetry of molecules of widely differing size can be made by calculating the diameter,  $R$ , which the molecule would have if it were spherical and unhydrated, and then computing the number of unit charges,  $z$ , which must be located at each end of a dipole of this length in order to give the observed dipole moment<sup>52</sup>. Values obtained in this way are all much smaller than half the total anionic and cationic charges, which is roughly the maximum possible value,  $z_{\max}$ . The ratio  $z/z_{\max}$  obtained for serum albumin is about 0.02, for hemoglobin 0.03, and for edestin 0.01. These molecules would accordingly have moments many times those reported here if they were very unsymmetrical electrically, and the measurements can be taken as proof of a fairly high degree of electrical symmetry.

The relaxation times observed for these proteins (with the possible exception of insulin) are of just the order of magnitude which would be expected from other values of molecular weight, hydration, and asymmetry. Since the molecular weight is the best known of these quantities, we have chosen to present this correlation in the form of a graph in which possible values of hydration (grams of water per gram of protein) and axial ratio ( $a/b$ ) are shown by various shaded areas. Such curves are given in Fig. 11 for nearly all the proteins for which many of the necessary data exist. The correlation of the various methods is rather good, in that a certain area can usually be chosen which will represent the "most probable" values of axial ratio and hydration. These graphs show very vividly that diffusion data combined with viscosity data cannot differentiate between various possible combinations of asymmetry and hydration. This is equally true of dielectric dispersion data if only one relaxation time has been observed. However, when the dielectric dispersion results lead to two relaxation times the axial ratio and hydration are uniquely determined, within limits depending on the probable errors in the relaxation times. Values from x-ray and from crystal density studies have been included in the preparation of Figure 11. It is, of course, true that a determination of hydration in a protein crystal gives no direct information concerning the hydration of the same protein in solution. Nevertheless, the agreement between these various methods for evaluating hydration and axial ratio yield the strongest support for the view that protein molecules appear to rotate as rigid units, which can be mathematically described as if they were ellipsoids of revolution of various axial ratio.

<sup>52</sup> We have used the equation

$$z = \mu / (eR) = 0.41 [(\Delta D_t / g)^{1/2} M^{1/6} / \bar{v}^{1/3}] = 0.41 [\delta_t^{1/2} / V^{1/3}]$$

where  $M$  is the molecular weight,  $(\Delta D_t / g)^{1/2}$  is the square root of the difference in dielectric constant divided by the partial specific volume of the protein, and  $\delta_t$  and  $V$  are the dielectric constant and the partial specific volume, respectively.

## Chapter 23

### The Solubility of Proteins

BY EDWIN J. COHN

#### I. THE SEPARATION AND CLASSIFICATION OF PROTEINS

Proteins differ from one another widely with respect to their solubility. On the basis of these differences the chemists of the eighteenth and nineteenth centuries learned to effect their extraction from plant and animal tissues. The solvents generally employed were: (1) water; (2) salt solutions; (3) alcohol-water mixtures; (4) acid or alkaline solutions. Often all these solvents were employed successively, generally in the above order. The order had been arrived at empirically, for two reasons. In the first place, many proteins are soluble in neutral salt solutions, although insoluble in water; and nearly all proteins dissolve in acid or alkaline solutions. Thus, if solvent (4) were employed before the others, no fractionation would be achieved, and the same would be true if solvent (2) were employed before solvent (1).

The most effective solvents often modify and denature proteins in the process of dissolving them. Thus, the second reason for the order in which the early chemists employed these solvents depended on their observation of denaturing effects. They noted that the state of most proteins was readily altered in acid and alkaline solutions as well as in alcohol-water mixtures.

The body of knowledge acquired by those concerned with the extraction and separation of proteins was reformulated at the beginning of this century into a classification of the proteins. Thus those soluble in water were called albumins, those soluble in salt solutions, globulins, those soluble in alcohol-water mixtures, prolamines, and those that were insoluble in all of these, but dissolved in acid and alkaline solutions, glutelins.

Although this classification is inadequate, and an international agreement was not reached respecting all the details of definitions, these four rough classes of proteins have a certain utility and physicochemical basis.

tions. Conversely, many proteins previously classified as conjugated have been freed of their foreign groups and isolated as pure proteins. The early classifications include a nomenclature for acid and alkali modified proteins or metaproteins as well as for the products of partial protein hydrolysis into proteoses, peptones and peptides. The multitude of the states which represent partial or complete modification and denaturation, whether or not accompanied by hydrolysis, is so great that little use can be made of these early attempts at standardization. Instead of classification we must substitute accurate physical chemical characterization of each modified state as well as of the native undenatured protein.

Of the solvents employed by earlier workers, few can be assumed to yield products resembling those that occur in nature, and then only if the conditions for the extractions and the subsequent fractionations are rigorously controlled and circumscribed. Indeed the conditions that should obtain during the purification of proteins should approach as closely as possible those of the native tissue, except that the temperature should be low. The physicochemical conditions which have been found most important in the separation and purification of proteins may for convenience be listed as follows.

(1) The temperature should be kept as low as possible and far lower than body temperature if the actions of the enzymes of tissue extracts are to be avoided. Even room temperature is too high for most stages in protein purification, which is best carried out as close to the freezing point of the solvent as possible, excepting when advantage is taken in the separation of proteins of the influence of temperature upon solubility. Change in temperature changes (a) the dissociation of the various free groups of proteins (heat of neutralization); (b) the forces between protein molecules in the solid state and between protein and solvent molecules in solution (heat of solution); and (c) the influence of dilute and concentrated salt solutions respectively as solvents and precipitants. Thus, in dilute salt solution, increase in temperature commonly increases solubility, whereas in concentrated salt solutions the reverse is often the case and increase in temperature may be an effective means of inducing crystallization from a concentrated salt solution saturated with respect to a protein component.

(2) The pH should be maintained as near that of the environment of the native protein as possible. Excessive acidity and alkalinity lead to modifications of many proteins, not all of which are completely reversible. The acidities or alkalinities which can be withstood, without irreversible changes being induced, vary greatly, however, from protein to protein. It is thus necessary to prove that every change in reaction to which a protein is

(3) The dielectric constant should be kept as high as possible. Although certain knowledge is still lacking regarding the dielectric constants of body fluids and tissues, it is known that many proteins yield solutions with dielectric constants higher than those of even so polar a solvent as water. Whether change in the solubility of proteins that have been partially purified and separated from the other proteins with which they existed in nature depends upon changes in the distinctive properties of their environment has not been proven, but it has repeatedly been observed that proteins are more stable in concentrated than in dilute protein solutions; that they are "protected" by a high concentration of the same or of foreign proteins.

(4) Organic solvents should be employed only with great caution, and when they are employed the temperature should be held as low as possible. The solvent action of organic molecules upon amino acids and peptides increases, as we have seen, with increase in the number of non-polar paraffin side chains or of non-polar benzene or pyrrolidine rings of the dipolar ion. Organic solvents, alcohol-water mixtures, acetone-water mixtures and propyleneglycol-water mixtures, are solvents only for those proteins which are rich in non-polar groups. Proteins of this kind are, as we have seen, termed prolamines.

Unless there is a predominance of non-polar side chains, organic solvents in small concentrations are to be considered as protein precipitants. High concentrations of non-polar solvents are always protein precipitants. This precipitating action may be considered to depend upon interactions both with the uncharged polar side chains of proteins and with the charged groups of proteins upon which their electric moments depend. The former effect may be expected to occur without deformation of the protein. The latter might be expected to result in such changes, and in experience one frequent result of the use of organic solvents, especially at ordinary temperatures, is denaturation of proteins.

The charged groups of proteins give rise to Coulomb forces between molecules. They also lead to electrical forces between the charged groups of the same molecule. Addition of organic solvents, like increase in temperature, leads to decrease in the dielectric constant of the solution and thus to increase in all electrostatic forces. The dangers inherent in the use of organic solvents are in this way related to the dangers in the use of high temperature. The latter increases the energy of the heat motion of all matter; the former decreases the dielectric constant and increases the electrostatic forces between the charged groups of the same as well as of neighboring protein molecules.

(5) Dilution of the protein solution should be avoided whenever possible.

cases proteins are observed to become grossly denatured and to separate as insoluble precipitates. In others, no change in solubility is observed, but measurements reveal that the molecular weight has diminished to a fraction of that characteristic of more concentrated solutions.

Possibly these effects depend upon the fact that all proteins are exposed in dilute aqueous solution to lower dielectric constants than would be characteristic of their more concentrated solutions. The empirical fact—advantage of which was first clearly taken by Sørensen, who concentrated his protein solutions in negative pressure dialyzers<sup>1</sup>—would appear to be that proteins may be retained without sensible change when they are preserved as very concentrated solutions, or in the solid state as precipitates or crystals.

(6) Dilute solutions of neutral salts increase the stability of proteins even in dilute solution. Moreover, they increase the solubility of proteins by diminishing their activity coefficients. This effect of ions on dipolar ions is comparable to that of interionic forces in so far as both depend upon Coulomb forces. Whereas interaction between ions increases with the ionic strength, and with the valence of the ions, interaction between ion and dipolar ion increases with the ionic strength and with the electric moments of the dipolar ion. Both types of interaction are diminished by increase of the temperature and of the dielectric constant, but the quantitative relation of these factors—temperature, dielectric constant and ionic strength—is different, as we shall see, for these two types of interactions.

Besides the interaction between ions and dipolar ions, which depends upon ionic strength of the electrolyte, there are those which depend upon the specific nature of the electrolytes. These are most pronounced in concentrated solutions of both the ions and the dipolar ions.

(7) Most proteins are precipitated from solution by sufficient concentrations of certain neutral salts. Salts with multivalent cations are generally the best solvents, and those with multivalent anions the best precipitants, for proteins. Although they dissolve proteins at low ionic strengths, phosphates, citrates and sulfates precipitate proteins in more concentrated solutions at far lower salt concentrations than do chlorides or nitrates. Among the sulfates the ammonium is far more soluble than the sodium salt. Many proteins cannot be precipitated by sodium sulfate because even its saturated solution is not sufficiently concentrated at room temperature. Ammonium sulfate is therefore employed in the separation and purification of many proteins, despite the inconvenience of employing a salt containing ammonia and the consequent complication of analytical procedures based upon the nitrogen determination. Although sodium

sulfate precipitates proteins in lower concentration than ammonium sulfate, its low solubility has demanded its being employed at relatively high temperatures where its solubility is higher<sup>2</sup>, whereas the lability of the protein molecule renders it desirable, as we have seen above, to carry out all procedures at the lowest possible temperature. Phosphates are also very good protein precipitants. Moreover phosphate buffers of known pH and ionic strength<sup>3, 4</sup> may be prepared and used even at low temperature. Thus proteins may be protected by these strong buffers\* from any change in pH.

Other salts have been used as protein precipitants, among them magnesium sulfate and various acetates and chlorides. These salts are less effective protein precipitants than those discussed above. Although they are often considered as specific precipitants for individual proteins, it remains to be demonstrated that their use is more specific than that of sulfates or phosphates in solutions of controlled temperature, pH and ionic strength.

In the past it has sometimes been supposed that certain salts were specific precipitants for certain proteins. This conception appears to depend upon the fact that some proteins are far more easily salted out than others. Thus among plasma proteins, fibrinogen has been considered precipitable by half saturation, and euglobulin by saturation with sodium chloride. Both of these proteins are also precipitable however by phosphates and sulfates and so are the more soluble serum proteins for which sodium chloride is not a precipitant. The separation of more soluble proteins demands the use of phosphates, sulfates or organic precipitants and the degree of separation effected depends less upon the nature of the reagent than upon the choice of optimal conditions of protein concentration, concentration of precipitant, pH and temperature.

## II. DIFFERENCES IN THE SOLUBILITY RELATIONS OF DIFFERENT PROTEINS

The precipitating action of a given salt on different proteins depends upon well known properties of the protein molecule. Some of these may be listed:

(1) A given protein is least soluble in the neighborhood of its isoelectric point in the presence, as well as in the absence, of neutral salts. The pH of minimum solubility varies with the nature and the concentration of the neutral salt. Except in very dilute salt solution this pH is generally different from the true isoelectric point of the protein.

<sup>2</sup> Howe, P. E., *J. Biol. Chem.*, 49, 93, 109 (1921); 57, 235, 241 (1923).

<sup>3</sup> Cohn, E. J., *J. Am. Chem. Soc.*, 49, 173 (1927).

<sup>4</sup> Green, A. A., *ibid.*, 55, 2331 (1933).

\* At neutral reactions potassium phosphate buffers are the more soluble and can conveniently be employed at concentrations up to 4 molar. The precipitating action of ammonium sulfate and potassium phosphate

(2) The solubility of proteins in the uncombined, salt-free state varies widely. This is true among those that separate in the crystalline and in the amorphous state. Thus, crystalline edestin, from hempseed<sup>5</sup>, and amorphous myosin, from rabbit or cow muscle<sup>6</sup>, are extremely insoluble in the absence of neutral salts. Hemoglobin (horse) is far more soluble<sup>7</sup>. Moreover, the hemoglobins of different species have quite different solubilities<sup>8</sup> and often different crystal forms<sup>9</sup> even though they may have the same molecular weight. Human hemoglobin is so soluble in the absence of neutral salts that it has generally been crystallized, like the albumins, from concentrated salt solutions<sup>10</sup>.

Egg albumin and serum albumin are far more acid molecules than the above, having isoelectric points slightly acid to pH 5. Casein and gelatin have isoelectric points in the same range. While casein has a solubility in the absence of salt of approximately 0.11 gram per liter<sup>11</sup>, egg and serum albumin can be prepared as solutions containing 20% or more protein in the neutral state. Indeed serum albumin can be said to be miscible with water in all proportions. Gelatin can be prepared either as an amorphous precipitate or as a gel.

The forces between molecules in the solid state as well as those between solvent and solute molecules determine solubility. Thus the greater the crystal lattice energy, the lower the solubility. Molecules whose charged groups and electric moments result in strong attractive forces between precipitated molecules will require far greater attractive forces with solvent molecules for solubility to result. The differences between the solubility of uncombined proteins will thus depend as much upon the forces in the solid state due to the nature and position of the charged groups as upon the interactions between dissolved molecules and solvent.

(3) Solubility in water or in another solvent will depend not only upon the forces between molecules in the solid state, but also between the electric moments and the specific groups of the solute and solvent molecules. Thus proteins rich in paraffin side chains and in pyrrolidine rings and poor in charged groups tend to dissolve in alcohol-water mixtures rather than in water, whereas those poor in non-polar groups, but rich in polar or electrically charged groups, tend to be precipitated by even small amounts of alcohol and acetone.

(4) Proteins that are insoluble in water but are rich in charged groups

<sup>5</sup> Osborne, T. B., and Harris, I. F., *Am. J. Physiol.*, **14**, 151 (1905).

<sup>6</sup> Edsall, J. T., *J. Biol. Chem.*, **89**, 289 (1930).

<sup>7</sup> Cohn, E. J., and ... *Physiol.*, **8**, 619 (1927).

<sup>8</sup> Landsteiner, K., ... *Gen. Physiol.*, **6**, 131 (1923-24).

<sup>9</sup> Reichert, E. T., ... *Differentiation and Specificity of Corresponding Proteins and Other Vital Substances in Relation to Biological Classification and Organic Evolution. The Crystallography of Hemoglobins.* Carnegie Inst. of Washington, Washington, D. C. (1909).

will become more soluble, as we have seen, in the presence of neutral salts or other dipolar ions, and this effect will be the greater the greater the dipole moment of the solute protein.

(5) Proteins that are rich in charged groups and soluble in water, such as the albumins, have thus far been found to have small dipole moments.

(6) Most proteins are more soluble when combined with acids or bases than in the neutral state. Protein salts obey many of the same rules as do other salts. Moderate increases in ionic strength increase their solubility; specific ions may however decrease their solubility with the appearance of new solid phases.

(7) Formation of salts between protein and protein as those between protein and ion may result either in compounds of greater or of lower solubility. Whereas in certain cases the higher solubility of mixtures of proteins than of pure proteins illustrates the former principle, insoluble protamine insulinate may be considered an example of the latter principle.

(8) The addition of base to an insoluble neutral protein increases solubility with formation of a soluble proteinate. Provided no insoluble proteinate is formed a strict proportionality should obtain between the base added and the increase in protein in solution.

(9) If a new saturating body is formed, however, the conditions that obtain are more complex and solubility will be equal to the sum of the solubility of the neutral protein and the insoluble proteinate plus the concentration of that alkali in the system, minus the alkali bound in the solid state times the equivalent combining weight of the soluble alkali proteinate. That is to say, the total alkali is divided between soluble and insoluble proteinate and solubility will depend upon the equivalent combining weight of both soluble and insoluble proteinates.

(10) Neutral salts often increase the solubility not only of neutral proteins but also of protein salts. Thus, the formation of an insoluble proteinate may, or may not, diminish the solvent action of neutral salt upon the protein. There are two conditions under which neutral salts instead of supplementing the solvent action of acid or alkali, by virtue of their ionic strength, have precipitating actions. The first of these follows the formation of a colloidal suspension when protein is dissolved with acid or alkali. Small amounts of neutral salts commonly discharge the electric forces holding the protein in a colloidal state as a result of which the system reverts to a condition more closely related to that predicted from the classical laws of solution. The second condition, under which neutral salts in relatively low concentration diminish the solubility of proteins combined with small amounts of acid or alkali, depends upon the phenomenon also in inorganic chemistry of the precipitating action



influence of small amounts of neutral salts on the precipitation of proteins from acid solution.

(11) Finally, in sufficiently concentrated salt solutions protein salts are precipitated as well as neutral proteins, though the salt concentrations necessary for precipitation are greater the more the protein is combined with acids or basis. Salting-out is thus generally maximal near the iso-electric point. It is also greater for otherwise comparable proteins the larger the protein and the larger its proportion of non-polar groups.

(12) The effect of temperature upon solubility in concentrated salt solutions is often the reverse of that in salt-free solutions, in solutions at low ionic strengths, or in ethanol-water mixtures. Under the former conditions, solubility may be decreased by increase in temperature, and crystallization would result, whereas in the latter, solubility increases with temperature much as with most other molecules.

### III. APPLICATION OF THE PHASE RULE TO THE SOLUBILITY RELATIONS OF PROTEINS

By JOHN T. EDSALL

In the study of proteins, solubility determinations are perhaps the most sensitive criteria of purity. The criteria most commonly applied in determining the purity of simpler compounds—namely, melting point and elementary composition—are of little or no value in the study of proteins, which decompose without melting and are all very similar in elementary composition. If a protein preparation shows constant solubility in a solvent of fixed composition, at constant pressure and temperature, even when the amount of undissolved protein in the system is varied within wide limits, this is very strong evidence that the preparation consists of a single protein component. So rigorous is this criterion of purity, indeed, that very few proteins have yet been clearly shown to fulfil it. Both the experimental methods and the underlying theory involve complexities which call for further discussion.

Among the earliest quantitative studies of the solubility of proteins were those of Osborne and Harris<sup>5</sup> on edestin, and those of Hardy<sup>12</sup> and of Mellanby<sup>13</sup> on serum globulin, in solutions of different salts. These solubility studies led to important advances in knowledge—notably to the formulation of the ionic strength rule by Mellanby—but neither serum globulin nor edestin behaved like a single protein component. Instead, as the total amount of protein in the system was increased the amount of protein dissolved by any given solvent also increased. For serum globulin, it was found that the amount of protein dissolved was almost directly pro-

portional to the amount added to a given volume of solvent. Such findings are not surprising in the light of our present knowledge, since serum globulin is now definitely known to be a mixture of several components, and edestin is a large protein molecule capable of dissociating into smaller ones.

In 1917 Sørensen<sup>14</sup> published his classic studies on egg albumin. He showed that this protein at fixed temperature, pressure, salt concentration and pH, behaved very nearly like a single component, showing a solubility almost independent of the amount of saturating body, in concentrated ammonium sulfate solutions. The increase in the amount of protein dissolved, however, with increase in the total protein of the system, was clearly well beyond the experimental error. One example of Sørensen's

TABLE 1. INFLUENCE OF TOTAL PROTEIN CONCENTRATION ON SOLUBILITY OF EGG ALBUMIN

Temperature of F : : : : : 18°. Filtrate of the fourth filtration contained 26.66 gm ammonium sulfate in 100 gm water; pH 4.92.

In total system Z	Protein hydrate in gm per 100 gm water In filtrate of 4th filtration S	$\Delta Z$	$\Delta S$	$D_t = \frac{\Delta S}{\Delta Z \cdot S_1}$
14.140	0.768			
9.419	0.726	4.721	0.042	0.0150
7.058	0.692	2.361	0.034	0.0243
4.703	0.654	2.355	0.038	0.0272
2.350	0.613	2.353	0.041	0.0294

$S_1$ , the value of  $S$  when  $Z = 1$ , was graphically determined as 0.593. The quantity " $D_t$ " in the last column is defined by Sørensen on the "dissociation tendency" of the protein.

From Sørensen, S. P. L., ref. 21.

numerous data is shown in Table 1. Similar phenomena, of a far more pronounced character, were revealed in Sørensen's laboratory by the study of the serum proteins<sup>15</sup>, gliadin<sup>16</sup> and casein<sup>17, 18, 19, 20</sup>. All these proteins were capable of fractionation into components of widely different solubility, although no one of the components could be shown to be a pure chemical individual by the solubility test. Consideration of these complex phenomena led Sørensen to the conception of proteins as "reversibly dissociable component systems"<sup>21</sup>. In his own words, "soluble

<sup>14</sup> Sørensen, S. P. L., and Høyrup, M., *Compt. rend. trav. lab. Carlsberg*, 12, 213 (1917).

<sup>15</sup> Sørensen, S. P. L., *Compt. rend. trav. lab. Carlsberg*, 15, No. 11 (1925); 16, Nos. 8 and 12 (1927); *J. Am. Chem. Soc.*, 47, 457 (1925); New York, (1924).

<sup>16</sup> Haugaard, G., *Compt. rend. trav. lab. Carlsberg*, 19, No. 9 (1930).

<sup>17</sup> Linderstrøm-Lund, L., *Compt. rend. trav. lab. Carlsberg*, 17, No. 9 (1928).

<sup>18</sup> Linderstrøm-Lund, L., *Compt. rend. trav. lab. Carlsberg*, 16, No. 1 (1925).

proteins consist of a series of complexes or components, reversibly combined, which makes their constitution expressible by the ordinary formula  $A_x B_y C_z \cdots A, B, C$  and so on each marking complete complexes, mainly polypeptides, yet in some cases also containing other groups, for example phosphorus groups, whereas the affixed indices  $x, y, z$ , and so on, mark the amount to which the indicated complex is present in the entire component system. Within each complex all the atoms and atomic groups are linked together by main valencies, whereas the various complexes in the whole component system are comparatively loosely and reversibly knit together by means of the residual valencies which each component must be assumed to possess, and the strength and nature of which must depend on the chemical composition of the component in question, as well as on its physical properties, above all on its dimensions and the resulting shape and surface. But all things, considered, the linkage between the components must be supposed to be comparatively slight and of such a nature that alterations in the composition of the solution (salt content, hydrogen-ion activity, alcohol content, temperature) may give rise to reversible dissociation of the involved component system and interchange of components between the same".\*

Sørensen interpreted the change in solubility of protein with change in the total protein content (dissolved and undissolved) as an expression of the dissociation tendency ( $Dt$ ) of the protein. An example of the calculation of this quantity is given in Table 1 for egg albumin. The dissociation tendency of this protein, according to this criterion, is low; that is, this proportion of egg albumin behaved quite nearly like a single component. For all proteins investigated, the dissociation tendency increased as the amount of protein dissolved diminished.

The later work of Sørensen and Sørensen<sup>22</sup> on the solubility of horse hemoglobin, carried out with the utmost care, confirmed the studies of Cohn and Prentiss<sup>7</sup> and of Green<sup>4, 10</sup> by showing that this protein has a very small dissociation tendency, and behaved very nearly like a single pure component.

The work of Sørensen, Linderstrøm-Lang, and their collaborators performed an immense service in revealing the complexity of many protein preparations that had previously often been regarded as pure components. Sørensen's treatment, however, did not lead to any method of predicting quantitatively the solubility of such a mixture. Later, Steinhardt<sup>23</sup> in a study of the solubility of crystalline pepsin at pH 2.8 in dilute KCl solutions, showed that the solubility steadily declined on equilibration with successive portions of the same solvent. The rate of the decline was found

to be proportional to the ratio of solid to the total volume of solvent employed, and was independent of the time taken to perform the whole series of equilibrations (provided a necessary minimum time was allowed for equilibrium to be reached at each step). Steinhardt expressed his results by the equation

$$S = A(1 - e^{-KV/A}) \quad (1)$$

where  $S$  is the total amount of protein dissolved by equilibration with the total volume,  $V$ , of solvent (whether the protein is treated with the solvent all at once or in many successive portions);  $A$  is the quantity of pepsin present at the beginning of the experiment;  $e$  is the base of natural logarithms, and  $K$  is a constant. Steinhardt found this equation to express his results very exactly, for pepsin from two different sources, in a variety of solvents; and he indicated that it should be applicable to some of the data of Sørensen on other proteins. Under the conditions of Steinhardt's experiments, pepsin contains significant amounts of non-protein nitrogen, both in the crystals and in the solution in equilibrium with them. Steinhardt showed that the experimental data could be explained on the assumption that the crystals form a solid solution, in which the activity of the pepsin is proportional to its mole fraction. The mole fraction of the pepsin is small, in the solid solution, although its weight fraction is large, since the impurities are of much lower molecular weight. If further the solubility of the non-pepsin components is low compared to that of the pepsin, then the relations to be expected theoretically should correspond with those described by equation 1. This equation is thus a special case, capable of further generalization for other types of solid solutions, and the extension of Steinhardt's analysis to other anomalous protein systems may yield valuable results.

The work of Northrop and his collaborators on crystalline enzymes<sup>24</sup> during the last ten years, however, has indicated that protein preparations can be obtained which rigorously fulfil the requirements of the phase rule for a pure single component. The phase rule<sup>25</sup> is commonly defined by the equation

$$P + F = C + 2 \quad (2)$$

Here  $P$  = number of phases,  $F$  = number of degrees of freedom, and  $C$  = number of components. The components selected to specify completely the composition of a given system may be chosen in more than one way, but the minimum number of components necessary to fix the composition of the system precisely is independent of this choice. A buffered solution of a single protein in salt and water must be regarded as a four-component

salt is ammonium sulfate, for instance, the four components may be chosen as water, protein, ammonium sulfate, and hydrogen ion (expressed as pH). They may alternatively be chosen as water, protein, ammonia, and sulfuric acid, since the pH and ammonium sulfate concentration can be specified by the amounts of  $\text{NH}_3$  and  $\text{H}_2\text{SO}_4$  in the system. The protein in the system consists of a large number of different ions and dipolar ions, all in equilibrium with one another (Chapter 20, p. 460). The protein, however, is a single component, since the entire equilibrium between these different forms is specified when the mass of total protein, and of the other components named, is fixed at a given temperature and pressure.

Temperature and pressure are always fixed in accurate solubility measurements. Under these circumstances it is apparent from equation 2 that  $F = C - P$ . Thus in this four-component system, containing two phases, if salt concentration and pH are fixed, the concentration of the dissolved protein is fixed. If only one phase is present, one degree of freedom still remains, and the amount of protein present in that phase may be varied at will, below the point at which a second phase appears. If an experiment is conducted by adding increasing amounts of protein to a solvent which is initially protein-free, the results may be most conveniently described by plotting the protein content of the solution as ordinate, against the total protein content of the system as abscissa. The resulting curve is a straight line of unit slope, up to the point at which the solution is saturated with the protein. Above this point the solid protein phase appears, and by the phase rule the composition of the solution is fixed, if the protein is a single pure component. In this region, therefore, the solubility curve should be a straight line parallel to the abscissa axis, the change from unit to zero slope being discontinuous at the saturation point of the solution with protein.

The first curves of this sort were obtained for chymotrypsinogen by Kunitz and Northrop<sup>26a, b</sup>. Recently Butler<sup>27</sup> has been able to carry out a further fractionation of chymotrypsinogen; the solubility curves obtained on some of his fractions are shown in Fig. 1. The fractions designated by Butler as A and A' give curves characteristic of solid solutions (see below), but fractions B, C and C' behave like pure single components, as nearly as can be judged from the data.

If two or more solid phases, of different solubility, are present, the curve will show breaks for several different points, each corresponding to the concentration at which saturation with one of the components has been achieved. When saturation with all components is attained, the composition of the solution is invariant, and the curve is horizontal above this

that he studied "seemed to vary with the buffer used". It thus becomes necessary to differentiate between the influence of neutral salts on the flocculation of edestin and on the insoluble compounds that this, and presumably other globulins, form with acids. It would seem as though Osborne's analytical procedure was less equivocal than the methods that have occasionally been applied to globulins without adequate consideration of their special chemical characteristics.

Edestin is possessed of but a small number of acid and basic groups that dissociate in the neighborhood of its isoelectric point. Osborne's measurements, interpreted on the basis of a molecular weight of 336,000, indicate that edestin contains 24 groups which can combine base in the neighborhood of the isoelectric point to form a soluble compound, and 24 which combine with acid to form the insoluble compound referred to above. A combination of acid with 24 other groups transforms the insoluble into a soluble hydrochloride<sup>14a</sup>. In all, therefore, 72 groups dissociate from pH 4.5 to 8.5.

Not only globulins but also albumins form insoluble salts with anions. Thus serum albumin is isoelectric near pH 4.8. It has generally been crystallized near this reaction from concentrated salt solutions; minimum solubility shifting somewhat to more acid reactions, the greater the ionic strength. Recently the observation was made by J. D. Ferry<sup>18</sup> that a fraction of serum albumin can also be crystallized from salt-free solutions near pH 4 as a somewhat insoluble sulfate—more soluble, however, at higher than at lower temperatures—and this phenomenon has been employed in the further fractionation and purification of the crystalline serum albumins<sup>19</sup>.

Specific interactions between protein cations and other anions, especially complex anions, have, of course, always been recognized and many of these lead to the formation of new saturating bodies. Many analytical and industrial processes, among them the combination of proteins with tannates, picrates, picrolonates and phosphotungstates, depend upon the formation of such insoluble protein salts. Combination of proteins with such anions often leads to changes in the proteins, however, which are not readily reversible.

A great many proteins are insoluble over a very wide pH range throughout which they combine with acids and bases to form a series of insoluble salts. The interpretation of the behavior of such protein solid phases demands a knowledge not only of the neutral molecule, but of the stoichiometric and electrochemical valences of the several insoluble compounds.

<sup>18</sup> Ferry, J. D., reported to the Division of Biological Chemistry at the 96th Meeting of the American Chemical Society, Milwaukee, Wis., September, 1938.

<sup>19</sup> McMeekin, T. L., *J. Am. Chem. Soc.*, 61, 2884 (1939).

point. Between the initial section of unit slope, and the final section of zero slope, lie one or more segments of the curve with slopes between zero and unity. In employing solubility as a criterion of purity it is especially

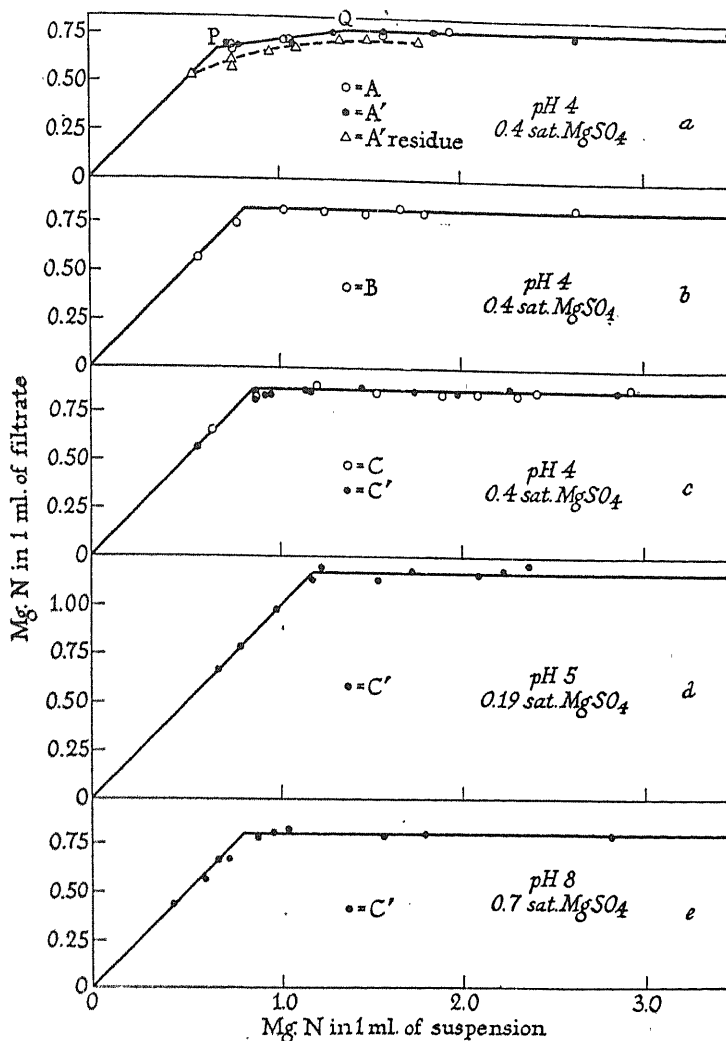


Figure 1. Solubility curves of chymotrypsinogen fractions in various solvents. From Butler, J. A. V., *J. Gen. Physiol.*, **24**, 189 (1940).

important to study the portion of the curve which corresponds to the "limit of solid phase at equilibrium" for in

In addition, the specific effects of both anions and cations upon the properties of such protein complexes must generally be considered<sup>20</sup>.

Proteins which form soluble sodium or potassium salts often form insoluble calcium acid salts. Thus casein passes into solution at 25°, as we have seen, with an equivalent weight of 2,100 grams casein per mole sodium hydroxide, or  $47 \times 10^{-5}$  mole of sodium hydroxide per gram of casein. In the presence of small amounts of calcium hydroxide, casein forms an insoluble calcium salt which, however, dissolves on the addition of further calcium hydroxide<sup>21</sup>. Casein dissolves as a soluble salt when combined with  $47 \times 10^{-5}$  mole of calcium hydroxide, or double the number of equivalents of base necessary to form a soluble sodium salt<sup>21a</sup>. That casein combines with calcium hydroxide in molecular and not in equivalent proportions has long been recognized<sup>21</sup> and the more complex solubility relations in systems which contain not only casein but casein calcium salts as saturating body have been studied<sup>22, 23</sup>. The influence of the calcium on the molecular weight of the casein in solution must also be taken into account<sup>23</sup>.

Certain of the serum proteins form insoluble calcium<sup>24</sup> and zinc salts, and this observation has been employed in the separation of serum globulins from one another<sup>25</sup>. The molecular weight of serum globulins as of casein has been noted to vary in the presence of calcium ions<sup>26</sup>. The isolation of various globulins as calcium and zinc salts must be expected to follow and to lead to the characterization of these complexes.

Insulin also forms an insoluble zinc salt, the amount of zinc in combination with the protein crystals increasing with increase in pH from 2 to 4 atoms per molecule. Dissociation of the zinc from zinc insulin compounds readily occurs, however, with liberation of zinc ions and of insulin containing smaller amounts of zinc<sup>27, 28, 28a</sup>.

Insulin also may be taken as an example of a protein anion forming insoluble salt with a protein cation. Protamine insulate is less soluble than insulin and the protamine<sup>29</sup> has been shown to displace zinc from its combination with insulin<sup>28</sup>. The very basic protamines have long been known to form insoluble caseinates, and insoluble protein complexes occur also between proteins that are less strong acids than insulin or casein and

<sup>20</sup> Steinhardt, J., *Anal. N. Y. Acad. Sci.*, **41**, 287 (1941). See Chapter XX, pp. 496-8.

<sup>21</sup> Van Slyke, L. L., and Hart, E. B., *Am. Chem. J.*, **33**, 461 (1905); Van Slyke, L. L., and Van Slyke, D. D., *ibid.*, **38**, 383 (1907).

<sup>21a</sup> Loeb, J., *J. Gen. Physiol.*, **3**, 547 (1920-21).

<sup>22</sup> Pertzoff, V., *J. Biol. Chem.*, **79**, 799 (1928).

<sup>23</sup> Philpot, F. J., and Philpot, J. St. L., *Proc. Roy. Soc. London*, **B 127**, 21 (1939).

<sup>24</sup> Mellanby, J., *J. Physiol.*, **33**, 338 (1905).

<sup>25</sup> Felton, L. D., *J. Immunol.*, **21**, 357 (1931).

<sup>26</sup> Svedberg, T., personal communication.

<sup>27</sup> Scott, D. A., and Fisher, A. M., *Biochem. J.*, **29**, 1048 (1935).

<sup>28</sup> Cohn, E. J., Ferry, J. D., Livingood, J. J., and Blanchard, M. H., *Science*, **90**, 183 (1939); *J. Am. Chem. Soc.*, **63**, 17 (1941).

<sup>28a</sup> Ferry, J. D., *The Collecting Net*, **14**, 154 (1939).

<sup>29</sup> Hagedorn, H. C., Jensen, B. N., Krarup, N. B., and Wodstrup, I., *J. Am. Med. Assoc.*, **106**, 177 (1936).



a solubility independent of the amount of solid present, even if the solid is composed of several different phases.

These considerations are illustrated by Kunitz's solubility measurements<sup>26b, 28</sup> on  $\alpha$  and  $\gamma$  chymotrypsin at pH 5.5, studied both separately and in combination (Fig. 2). Neither protein is quite pure, although both are nearly so; but the mixture of the two shows sharp breaks at B and C

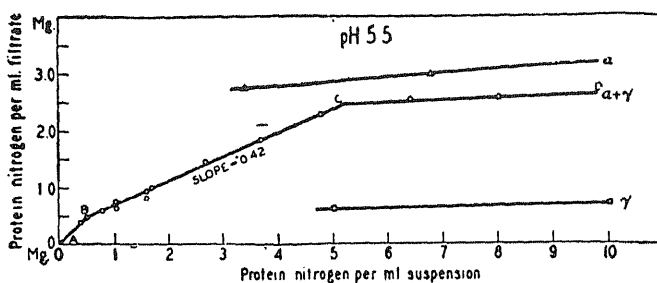


Figure 2

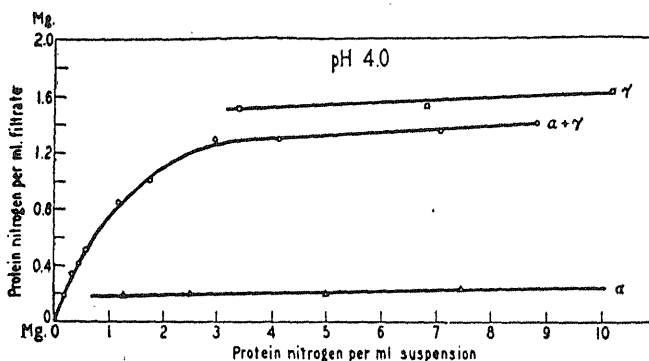


Figure 3

Figures 2 and 3. Solubility curves of artificial mixtures of crystals of alpha and gamma chymotrypsin (40 per cent alpha + 60 per cent gamma) in 0.4 saturated ammonium sulfate of pH 4.0 and 5.5 at 10°C. in the presence of increasing quantities of solid phase. At pH 4.0 the curve is of solid solution type while at pH 5.5 the curve corresponds to the theoretical curve of a mixture of two independent solid phases. Slope of BC, measured = 0.42, calculated = 0.43. From Kunitz, M., *J. Gen. Physiol.*, **22**, 207 (1938-39).

(Fig. 2, middle curve). If only the portion CD of this curve had been studied, it might have been taken as evidence for constant solubility by less careful and critical workers.

Systems containing two or more protein components may give solubility curves which, when plotted in this fashion, are exactly like those of a single

less strong bases than the histamines and protamines. Stoichiometric combinations between protein ions would also appear to lie at the basis of the so-called precipitin reaction and thus play a large role in immunity as well as in other aspects of protein chemistry<sup>30, 31, 32</sup>.

## VII. SOLUBILITY IN CONCENTRATED SALT SOLUTIONS: SALTING-OUT

The precipitation of proteins by neutral salts has been employed as a method for their separation, purification, characterization and occasionally classification<sup>8, 33</sup> ever since the procedure was first employed in the middle of the last century by Panum<sup>34</sup>, Virchow<sup>35</sup> and Claude Bernard<sup>36</sup>. Hof-

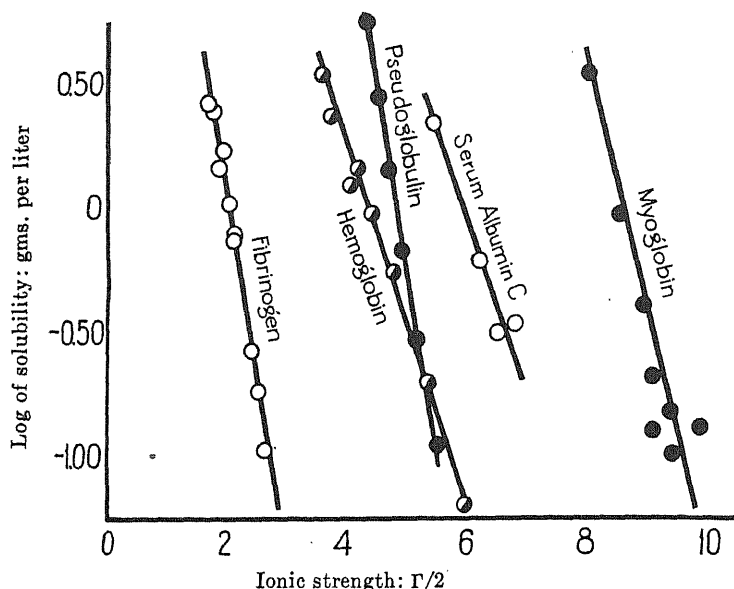


Figure 5. The solubility of proteins in ammonium sulfate solutions

meister demonstrated that salting-out depended upon the character of the neutral salt as well as of the protein, and his studies of solubility in concentrated salt solution have since been supplemented by those of Chick and Martin<sup>37</sup>, Sørensen and his co-workers<sup>38</sup> and by various other studies<sup>8, 33, 39, 40</sup>. Certain of the results in concentrated ammonium sulfate

<sup>30</sup> Heidelberger, M., *Cold Spring Harbor Symp. Quant. Biol.*, 6, 369 (1938); *Chem. Rev.*, 24, 323 (1939).

<sup>31</sup> Marrack, J. R., "The Chemistry of Antigens and Antibodies," 2nd ed., London, H. M. Stationery Office (1938).

<sup>32</sup> Pauling, L., *J. Am. Chem. Soc.*, 62, 2043 (1940).

<sup>33</sup> Green, A. A., *J. Biol. Chem.*, 95, 47 (1932).

<sup>34</sup> Panum, F., *Virchow's Arch. path. Anat.*, 4, 419 (1852).

<sup>35</sup> Virchow, R., *ibid.*, 6, 572 (1854).

<sup>36</sup> Bernard, C., in Robin, C., and Verdeil, F., *Traité de chimie anatomique*, Paris, 3, p. 299 (1853).

<sup>37</sup> Chick, H., and Martin, C. J., *Biochem. J.*, 7, 380 (1913).

<sup>38</sup> Sørensen, S. P. L., and his co-workers, *M. Chem. rend. trav. lab. Carlsberg*, 12, 213 (1917).

<sup>39</sup> Morgan, V. E., *J. Biol. Chem.*, 112, 537 (1935).

<sup>40</sup> Florkin, M., *ibid.*, 87, 629 (1930).

pure protein. Such a situation may arise in two ways<sup>27, 29</sup>: (a) the solid, from the moment of its first appearance, consists of two distinct phases. This can occur only if the two proteins are present in the system in amounts proportional to their solubilities. (b) Only one solid phase is present, but this phase is a solid solution of two proteins which dissolve in the same proportion in which they are present in the solid. In either case, if the different proteins differ significantly in structure, and especially in their acidic and basic properties, it is to be expected that their relative solubilities should not be the same in all solvents. It is therefore important, if a rigorous test of purity is sought, to repeat the solubility curve determination in several different solvents, of different pH and different salt concentration. The choice of solvents is, of course, restricted in practice. The solvent must not denature the protein, and the solubility of the protein in any solvent chosen must be large enough for accurate measurement, and small enough to permit making many determinations on the protein sample available.

In general, if the solid phase is a solid solution, the components do not dissolve in the same proportions in which they are present in the solid. Then the curve shows no sharp breaks, but passes fairly smoothly from the initial portion of unit slope to the final portion of zero or nearly zero slope. This situation is illustrated by Fig. 3, showing Kunitz's data for  $\alpha$  and  $\gamma$  chymotrypsin at pH 4.0. It is interesting that these two proteins form a solid solution at this pH, but are present as two distinct solid phases at pH 5.5 (Fig. 2). The theory of solid solutions, as applied to protein solubility, has been discussed in great detail by Northrop and Kunitz<sup>29</sup> and by Butler<sup>27</sup>.

Recently, Herriott, Desreux and Northrop<sup>30</sup> have shown that the solubility curves of ordinary crystalline pepsin preparations indicate the presence of more than one protein. They have separated one of the proteins involved, and showed it to have constant enzymatic activity and constant solubility in several solvents. In certain solvents—notably those containing acetate buffer—the solubility was found to be markedly affected by the amount of non-protein nitrogen decomposition products present, while in other solvents this was not the case. The most favorable conditions for obtaining constant solubility were found to involve pH values generally between 4 and 5, where the rate of formation of non-protein decomposition products was much slower than in the more acid solvents employed by Steinhardt<sup>23</sup> and other investigators.

The most effective method of fractionation is to equilibrate the protein with a solvent in which a considerable amount dissolves, but much is left

solutions are graphically represented in Fig. 5 in which the logarithm of the solubility is plotted as ordinate and the ionic strength,  $\Gamma/2$ , as abscissa. The linear relation that then obtains was demonstrated in 1925<sup>41</sup>, and its significance discussed.

Salting-out is characteristic not only of proteins but of gases, of uncharged molecules and, in sufficiently concentrated solutions, of electrolytes, and is given by the equation:

$$\log S = \beta' - K'_s(\Gamma/2) \quad (9)$$

$K'_s$ , the salting-out constant, gives the slope,  $\beta'$  the extrapolated intercept of the linear portion of the curve on the ordinate axis. If the solubility and the ionic strength, are expressed as gram per thousand grams of water, a similar logarithmic equation also applies.

$$\log S = \beta - K'_s I/2 \quad (9a)$$

or, if solubility is expressed as mole fraction of protein,

$$\log N = \beta - K'_s \Gamma/2 \quad (9b)$$

The fit of the equations to the data is almost equally good in all these expressions<sup>42</sup>, but the values of the constants involved ( $\beta$  or  $\beta'$ , and  $K_s$  or  $K'_s$ ) will naturally be somewhat different according to the form of expression which is used.

These equations describe the solubility of proteins only over a limited range in concentrated salt solutions where the solubility is sufficiently small. Indeed they hold only when the salting-out effect is large in comparison with the solvent action of neutral salts; the apparent constant,  $K'_s$ , representing a summation of both effects. The exactly comparable behavior of cystine in ammonium sulfate thus led to an apparent value of  $K'_s$  of 0.044 (Table 2, Chapter 11), whereas the true salting-out constant,  $K_s$ , was estimated to be far greater, namely, 0.14 (Chapter 11). Before considering the interaction of ions with proteins as dipolar ions, we shall therefore first consider the simpler conditions that appear to obtain in very concentrated salt solutions.

#### Values of $K'_s$

The salting-out constant,  $K'_s$ , for a given protein and a given salt, within the experimental error, is independent both of temperature and pH, whereas  $\beta'$  is markedly influenced by pH and temperature. Values of  $K'_s$  may thus be employed to characterize the interaction of a particular protein with a concentrated solution of a given salt. Values of  $K'_s$  for a number of different proteins in different salt solutions are given in Table 6; a few values for amino acids are also given for comparison.

<sup>41</sup> Cohn, E. J., *Physiol. Rev.*, 5, 349 (1935).

<sup>42</sup> Åkerlöf, G., and Thomas, H. C., *J. Am. Chem. Soc.*, 56, 593 (1934)

undissolved. The soluble portion is taken off, precipitated, and then reextracted with the same solvent used in the first extraction. The pepsin fraction which dissolves on the second extraction has a solubility curve such as would be expected for a single component<sup>30</sup>. Further fractionations may of course be undertaken, either on the less soluble or on the more soluble portions from a given extraction. This method of repeated extraction and precipitation with the same solvent has proved very effective also in the purification of diphtheria antitoxin<sup>31</sup>. It is effective in separating two or more proteins, whether they are present as separate solid phases, or as a solid solution<sup>32</sup>.

Other proteins which behave nearly like pure components, by Northrop's method, are trypsin<sup>33</sup>, salmon pepsin<sup>34</sup>, ribonuclease<sup>35</sup>, pituitary luteinizing hormone<sup>36</sup>, and lactogenic hormone<sup>37</sup>. The studies on these proteins, however, are less extensive and detailed than those on pepsin and chymotrypsin.

Certain other proteins behave, by the test of solubility, very nearly like pure single components. Among these are egg albumin and horse carboxyhemoglobin, already mentioned as having been studied by Sørensen. Others are human hemoglobin<sup>38</sup>, myoglobin<sup>39</sup>, fibrinogen<sup>40</sup>, lactoglobulin<sup>41</sup>, and a fraction of crystalline serum albumin<sup>42</sup>.

### The Thermodynamic Activity of the Protein in the Crystalline Phase

In Chapter 24, the effect of pH, of the dielectric constant of the medium, and of other factors on protein solubility will be considered. If two different solvents are in equilibrium with the same solid phase, then the activity of any component in the two solutions is fixed, and is equal to the activity of that component in the solid phase. From the equal activity of this component in the two solutions, it immediately follows that the ratio of the activity coefficients in the two solutions is inversely as the ratio of the solubilities.

It has, however, been shown<sup>43</sup> that the density of crystals of many, if not all, proteins is a function of the composition of the medium with which they are in equilibrium. The crystals are composed not only of protein, but of varying amounts of water, and probably of other components of the solvent. The alteration in the size of the unit cell in protein crystals,

<sup>30</sup> Northrop, J. H., *J. Gen. Physiol.*, **25**, 465 (1941-42).

<sup>31</sup> Herriott, R. M., *Chem. Rev.*, **30**, 413 (1942).

<sup>32</sup> Kunitz, M., *J. Gen. Physiol.*, **21**, 601 (1938).

<sup>33</sup> Norris, E. R., and Elan, D. W., *J. Biol. Chem.*, **134**, 443 (1940).

<sup>34</sup> Kunitz, M., *J. Gen. Physiol.*, **24**, 15 (1940).

<sup>35</sup> Shedlovsky, T., Rothen, A., Greep, R. O., van Dyke, H. B., and Chow, B. F., *Science*, **92**, 178 (1940).

<sup>36</sup> Li, C. H., Lyons, W. R., and Pomeroy, H. M., *J. Gen. Physiol.*, **24**, 303 (1940).

<sup>37</sup> Green, A. A., Cohn, E. J., and Pomeroy, H. M., *J. Biol. Chem.*, **109**, 631 (1935).

<sup>38</sup> Morgan, V. E., *J. Biol. Chem.*, **112**, 177 (1935).

as determined from x-ray diffraction (Chapter 14) on drying or as a result of less drastic changes in the solvent, provides even more direct evidence that these materials are an integral part of the protein crystal. The effect of such variations in composition on the thermodynamic activity of the protein component of the crystal is not known. If such knowledge is later obtained, it will place the thermodynamic interpretation of protein solubility studies on a firmer foundation. Even at present it is possible, as has been shown by Joseph<sup>44</sup> in a series of electromotive force studies, to measure the effect of a protein on the activity of a salt, and thereby to infer, by a simple thermodynamic relation,\* the effect of the salt on the activity of the protein. The calculated effects are of the same order of magnitude as those deduced from solubility studies, taking the activity of the protein in the crystalline phase as a constant. It appears probable, therefore, that the interpretation of protein solubility in terms of activity coefficients, as developed in Chapter 24, is a correct first approximation.

<sup>44</sup> Joseph, N. R., *J. Biol. Chem.*, 116, 353 (1936); 126, 389, (1938). See Chapter 24, p. 619.

\* See Chapter 11, Equation 16.

In general the values of  $K'_s$  for the proteins in Table 6 are ten to fifty times as large as for the amino acids (with the exception of tyrosine, whose salting out constant is reported to be remarkably large). In a logarithmic equation, differences of this magnitude in  $K'_s$  produce very great differences in the effect of salts on solubility. Thus the solubility of leucine if reduced approximately 11 per cent by 0.5*N* sodium chloride, while the solubility of fibrinogen is reduced by nearly 1000 per cent by a change in the ionic strength of a phosphate buffer solution from 1.5 to 2.0. Values of  $K'_s$  for different proteins and the same salt, or for different salts and the same protein, have not been found to differ by a factor of more than

TABLE 6. VALUES OF THE SALTING OUT CONSTANTS,  $K'_s$ , FOR AMINO ACIDS AND PROTEINS

Substance	NaCl	MgSO <sub>4</sub>	(NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub>	Na <sub>2</sub> SO <sub>4</sub>	Phosphate
Cystine*			0.05		
$\alpha$ -Aminobutyric acid (1).....	0.04				
Leucine (2).....	0.09				
Tyrosine (3).....	0.31				
Lactoglobulin (4).....				0.63	
Hemoglobin (horse) (5).....		0.33	0.71	0.76	1.00
Hemoglobin (man) (6).....					2.00
Myoglobin (7).....			0.94		
Egg albumin (8, 9).....			1.22		
Fibrinogen (10).....	1.07		1.46		2.16

Some of these values are slightly revised from earlier estimates (9, 11), where certain data have been rechecked and recalculated.

\* From the data shown in Fig. 4, p. 243, Chapter 11.

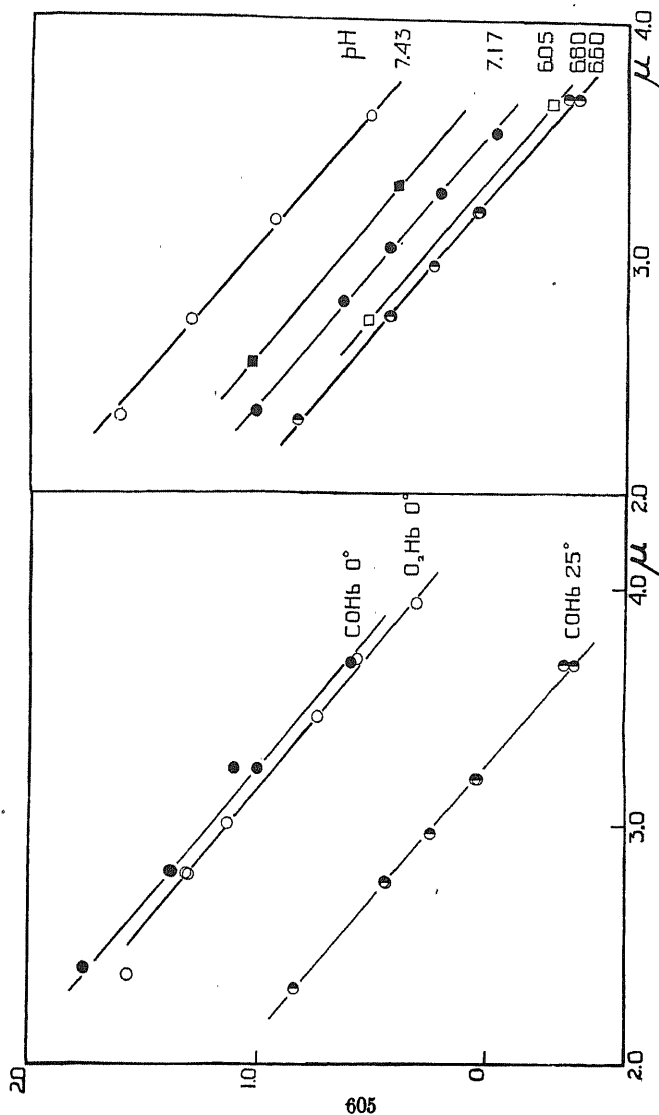
- (1) Joseph, N. R., *J. Biol. Chem.*, **111**, 479, 489 (1935).
- (2) Pfeiffer, P., and Würzler, J., *Z. physiol. Chem.*, **97**, 128 (1916).
- (3) Euler, H. v., *J. Biol. Chem.*, **110**, 113 (1924).
- (4) Palmer, A. H., *J. Biol. Chem.*, **110**, 34; Pederson, K. O., *Biochem. J.*, **30**, 961 (1936).
- (5) Green, A. A., *J. Biol. Chem.*, **97**, 128 (1916).
- (6) Green, A. A., *J. Biol. Chem.*, **97**, 128 (1916); M. H., *ibid.*, **109**, 631 (1935).
- (7) Morgan, V. E., *J. Biol. Chem.*, **112**, 112 (1917).
- (8) Sørensen, S. P., *Ann. N. Y. Acad. Sci.*, **12**, 213 (1917).
- (9) Cohn, E. J., *Plasma Protein Fractionation*, 2nd ed., New York, 1946.
- (10) Florkin, M., *J. Biol. Chem.*, **110**, 113 (1924).
- (11) Cohn, E. J., *Ann. Rev. Biochem.*, **4**, 83 (1935).

about two. Thus the value of  $K'_s$  for fibrinogen is slightly more than twice as great as for horse hemoglobin, whether potassium phosphate buffers or ammonium sulfate be used as the precipitating agent; and the efficiency of potassium phosphate as a precipitating agent for fibrinogen (as measured by the value of  $K'_s$ ) is almost exactly twice as great as that of sodium chloride.

For proteins as for simpler molecules, salts containing multivalent anions—phosphates, sulfates, and citrates\*—are the most efficient salting-out agents. Sodium sulfate is definitely more effective than ammonium sulfate, but the higher solubility of the latter salt often renders its use more convenient. Sodium and potassium chloride salt out fibrinogen and

\* The normal carbonates, which in general have a powerful salting-out action, cannot often be employed as protein precipitants on account of their high alkalinity.

Figure 6. The solubility of hemoglobin in concentrated phosphate buffers of varying temperature and pH. From Green, A. A., *J. Biol. Chem.*, **93**, 507 (1931).  
Abscissa: Ionic strength ( $I/2$ );  
Ordinate:  $\log S/S_0$ .





## Chapter 24

### Interactions of Proteins with Ions and Dipolar Ions \*

BY EDWIN J. COHN AND JOHN D. FERRY

#### I. SOLUBILITY AND THE HEAT OF SOLUTION

The more precise the criteria that are employed in demonstrating that proteins are chemical individuals, and the more nearly protein preparations satisfy these criteria, the more exact will be the physical chemical laws that are deduced to describe systems containing proteins. Crystalline proteins are often more readily purified than amorphous products, but crystallizability is not an indication that a single chemical substance has been separated. Thus repeatedly recrystallized serum albumin preparations have been shown by Sørensen to exhibit profound dependence on the amount of protein in the system; thus not satisfying the phase rule criterion for a chemical individual. On the other hand, casein, an amorphous protein, can be prepared in such a state of purity as to exhibit solubility behavior far closer to that demanded by the phase rule.

✓ Thus if our systems contain only one protein and water and the temperature be maintained constant, the solubility of the protein should also be constant. Some early measurements upon casein illustrate the extent to which this criterion was satisfied. The amount of protein in each system at the beginning of the experiment varied twenty-fold. After the precipitate had been adequately triturated with successive aliquots of solvent, in this case water at 25°, solubility approached a constant value independent of the amount of protein in the system and of the number of times it was equilibrated with solvent. Had the preparation been completely pure, solubility should have been the same in the first addition of solvent as in the last. This was not achieved since the casein could be washed free of impurities of higher solubility by repeated trituration. None the less, the solubility of casein can be taken from these experiments at close to 18 mg of casein nitrogen per liter, or 0.11 gram of casein per liter at this temperature (Table 1).

The solubility of a protein in any solvent must, as we have seen, be a function of the temperature. Solubility of casein in water has been studied

TABLE 1. SOLUBILITY OF CASEIN V

T. = 25.0° ± 0.1°C.

		Protein nitrogen suspended in 100 cc. H <sub>2</sub> O at beginning of experiment				
Experiment No.	Date 1922	mg. 8.55	mg. 17.10	mg. 42.75	mg. 85.50	mg. 171.0
		Protein nitrogen in 25 cc. filtrate				
		mg.	mg.	mg.	mg.	mg.
39	Apr. 3	0.47	0.42	0.47	0.48	0.72
		0.42	0.37	0.47	0.50	0.92
		0.42	0.47	0.50	0.47	0.72
	Apr. 11	0.50	0.40	0.50	0.45	0.50
		0.35	0.36	0.45	0.45	0.51
		0.40	0.55	0.45		0.50
	Apr. 14			0.40	0.40	0.40
				0.40	0.50	0.40
				0.40	0.45	0.45
	Average.....	0.43	0.43	0.45	0.46	0.46

TABLE 2. SOLUBILITIES AND HEATS OF SOLUTIONS OF VARIOUS PROTEINS

Protein	Solvent	Temperature (°C.)	Solubility (gm. per liter) S	Solubility of neutral molecule S <sub>n</sub>	Calculated heat of solution* ΔH
Zein (1)	Water	25	0.054		
Casein (2, 3)	Water	5	0.043		7800
		25	0.11	0.09	
Insulin† (4)	Water	5	0.009		positive
Lactoglobulin (5)	Water	5	0.35		
		25	0.58		4200
		0.25M glycine	5	2.6	
		25	3.3		2000
		0.50M glycine	5	6.8	
Egg albumin (6)	2.14M ammonium sulfate	25	9.1		2400
		0	3.18		negative
		12	2.09		
		20	1.81		positive
Oxyhemoglobin (7, 8, 9)	Water, pH 6.6	29	2.24		
		0	11.2		
Carboxyhemoglobin (10)	1.267M phosphate buffer, pH 6.6†	17.0	11.0		positive
		0	22.7		
Serum albumin sulfate (horse) (11)	Water	25	2.58		-18000
		5	120		
		25	80		-3300

\*Calculated by the equation  $\Delta H = \frac{R \ln (S_2/S_1)}{(1/T_1) - (1/T_2)}$ . This equation cannot be expected to hold for such concentrated solutions as for instance those of serum albumin sulfate. The calculations are none the less made since they show the sign and order of magnitude.

† Insulin sulfate resembles serum albumin sulfate (personal communication of J. D. Ferry) in that it has a negative heat of solution.

‡ The heat of solution is independent of phosphate concentration from 0.9 to 1.8M, over which range solubility is primarily determined by the salting-out effect, and the solubility diminishes about ten-fold.

(1) Cohn, E. J., Berggren, R. E. L., and Hendry, J. L., *J. Gen. Physiol.*, 7, 81 (1924).

(2) Pertzoff, V., *J. Gen. Physiol.*, 10, 961 (1927).

(3) Cohn, E. J., *J. Gen. Physiol.*, 4, 697 (1922); Cohn, E. J., and Hendry, J. L., *ibid.*, 5, 521 (1923).

also myosin<sup>14</sup> although at higher salt concentration than sulfate; but their salting-out effect on hemoglobins and albumins does not lead to a marked decrease in protein solubility in their saturated solutions. The salting-out tendency of calcium chloride, and presumably of other salts containing bivalent cations, is apparently still smaller for proteins, as it is for amino acids (Chapter 11).

### $\beta$ as a Function of pH.

In contrast to the constancy of  $K'_s$  for a given protein and a given salt, the variation of  $\beta$  with pH and temperature is very marked.  $\beta$  is the logarithm of a purely hypothetical protein solubility obtained by backward

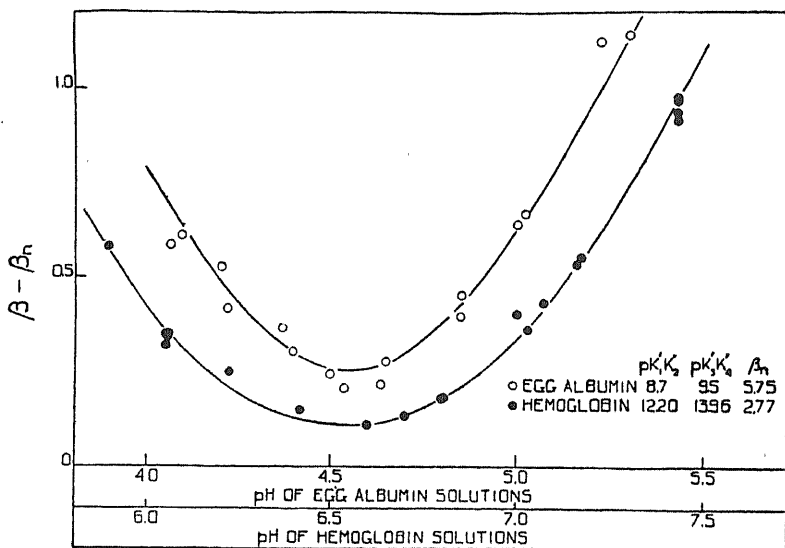


Figure 7. The solubility of hemoglobin and of egg albumin in concentrated salt solutions of varying pH. From Green, A. A., *J. Biol. Chem.*, **93**, 517 (1931).

extrapolation of the linear fraction of curves such as those in Fig. 5. Its variation with pH reflects the influence of the ionization of the protein upon solubility in concentrated salt solutions. Plots of  $\log S$  against ionic strength at various values of pH yield a series of parallel lines, as shown in Fig. 6, taken from the work of Green<sup>8</sup> on hemoglobin. At a given ionic strength, solubility as a function of pH may be described by equation 6 for both hemoglobin and egg albumin. Fig. 7 shows the variation of  $\beta$  with pH in concentrated salt solutions for these two proteins. The curves passing through the experimental points are calculated from equation 4 or 6, using suitable values of  $S_n$  and of the constants  $K'_1K'_2$  and  $K'_3K'_4$  (Table 5). Since equation 9 holds for these systems, and  $K'_s$  is independent

at more than one temperature<sup>1</sup> and reported to be 7 mg of casein nitrogen per liter or close to 0.043 gram casein in one liter at 5°. The heat of solution of uncombined casein in water may be calculated from these solubility measurements at 5 and 25° to be 7800 calories from the relation

$$\Delta H = \frac{R \ln (S_2/S_1)}{(1/T_1) - (1/T_2)}$$

Lactoglobulin and several other proteins have been studied in both water and other solvents, and these results also are represented in Table 2. Whereas more accurate measurements of these and other proteins in these and other solvents may subsequently be reported, the results at present available suffice to illustrate the magnitude of these quantities.

## II. SOLUBILITY AND EQUIVALENT COMBINING WEIGHTS

Solubility should be independent of the amount of the saturating body, not only in water but in any other solvent of fixed composition and tem-

TABLE 3. SOLUBILITY OF CASEIN IN SYSTEMS CONTAINING THE PROTEIN AND SMALL AMOUNTS OF SODIUM HYDROXIDE

NaOH added to casein $N \times 10^{-3}$	Hydrogen ion concentration $N \times 10^{-3}$	Protein N in 100 cc. Protein N dissolved in 10 cc.			Average protein N dissolved in 10 cc. mg.
		80 mg. mg.	160 mg. mg.	320 mg. mg.	
0.5	0.50	0.29			0.29
1.0	0.34	0.43	0.43		0.43
1.5	0.28	0.57			0.57
2.0	0.24	0.74	0.73	0.75	0.74
2.5	0.21	0.91			0.91
3.0	0.20		0.98		0.98
4.0	0.17	1.35		1.34	1.34
5.0	0.15	1.77	1.73		1.75
6.0	0.14			1.89	1.89

From Cohn, E. J., and Hendry, J. L., *J. Gen. Physiol.*, **5**, 521 (1923).

perature. Many proteins, as we have seen, have low solubilities in the neighborhood of their isoelectric points and may there be purified and often crystallized. Those that are globulins are rendered more soluble in the neighborhood of their isoelectric points by electrostatic interaction with ions and sometimes with dipolar ions. Most proteins, however, are more soluble when combined with acids or bases than when uncombined. The increase in solubility of proteins in systems containing acids or bases is, moreover, often stoichiometric in nature. This is illustrated in Table 3 by some early measurements upon the solubility of casein in the presence of varying amounts of sodium hydroxide at 25°.

In these experiments the amount of protein in systems containing the same amount of base was varied four-fold without measurably affecting

of pH, these curves also represent the variation of the logarithm of the actual protein solubility ( $\log S$ ) at constant ionic strength, as a function of pH.\*

Sørensen<sup>43</sup> has shown that the solubility of horse carboxyhemoglobin in concentrated ammonium sulfate passes through two minima, one near the isoelectric point at pH 6.6, the other in acid solutions at pH 5.4. He interprets this second minimum as due to the formation of an insoluble hemoglobin sulfate, containing 12 or 13 sulfate ions per mole of hemoglobin. The insolubility of such protein salts, especially in the presence of neutral salts, has already been discussed.

#### $\beta'$ as a Function of Temperature.

The value of  $\beta'$  is very sensitive to temperature, as well as to pH. A series of solubility measurements, at various temperatures and ionic strengths, plotted logarithmically in the salting-out range, forms a series of parallel straight lines (Fig. 6). Horse carboxyhemoglobin is approximately ten times as soluble at 0° as at 25° under these conditions, although at low salt concentrations its solubility is increased by rise of temperature. Thus instead of a positive heat of solution that of hemoglobin is negative and equal to -18000 calories in concentrated phosphate solutions (Table 2). In myosin<sup>14</sup> the decrease in solubility with rise of temperature, in the salting-out range, is even more marked; a solution which is highly soluble at 0° may be almost completely precipitated at 25°. The effect of temperature on the solubility of egg albumin, on the other hand, is relatively slight; the measurements of Sørensen (38) indicating that solubility passes through a slight minimum at about 20° (see 8 and Table 2). Although some proteins are more soluble, others are much less soluble under given conditions at low than at high temperature.

The varying response to pH and temperature of different proteins proves of practical importance in their separation. Constituents of a complex protein mixture which are very similar in solubility at a given pH and temperature may be affected very differently by the variation of one or the other of these factors.

A vast number of investigations of the fractionation of the serum proteins by salting out have been made; of those in relatively recent years the work of Howe<sup>6</sup> may be particularly mentioned. Still more recently Butler and his collaborators<sup>44, 45</sup>, Kydd<sup>46</sup>, and Jameson<sup>47, 48</sup> have analyzed the salting

\* The ordinate in Fig. 7 is not  $\beta$ , but  $\beta$  minus a constant,  $\beta_n$ . This is defined by the relation:

$$S/S_n = \text{antilog } \beta / \text{antilog } \beta_n$$

The method of calculating  $\beta_n$  is explained in detail by Green<sup>8</sup>. Its absolute value is immaterial in the present connection, since it does not affect in any way the change in  $\log S$  with pH, shown in Fig. 7.

<sup>43</sup> Sørensen, S. P. L., and Sørensen, M., *Compt. rend. trav. lab. Carlsberg*, 19, No. 11 (1933); *Biochem. Z.*, 258, 16 (1933).

<sup>44</sup> Butler, A. M., and H. A. H., *J. Biol. Chem.*, 99, 173 (1932).

<sup>45</sup> Butler, A. M., Blatt, J. H., and H. A. H., *ibid.*, 109, 755 (1935).

<sup>46</sup> Kydd, D. M., *ibid.*, 109, 100 (1935).

<sup>47</sup> Jameson, E., *Cold Spring Harbor Symp. Quant. Biol.*, 6, 331 (1938).

<sup>48</sup> Jameson, E., and Roberts, D. B., *J. Gen. Physiol.*, 21, 249 (1937).

but slightly with the amount of casein in the solid phase, and it was therefore concluded that true heterogeneous equilibria had been attained in these systems.

When the solubility measurements upon casein in systems containing small amounts of sodium hydroxide (Table 3) are plotted as ordinates and the base as abscissae, the points fall on a straight line. Accordingly, one must conclude that each mole of base carried an equivalent weight of casein into solution. The equivalent combining weight of casein given by the slope of the straight line in Fig. 1, as well as by measurements in systems containing far larger amounts of sodium hydroxide<sup>2</sup>, was estimated

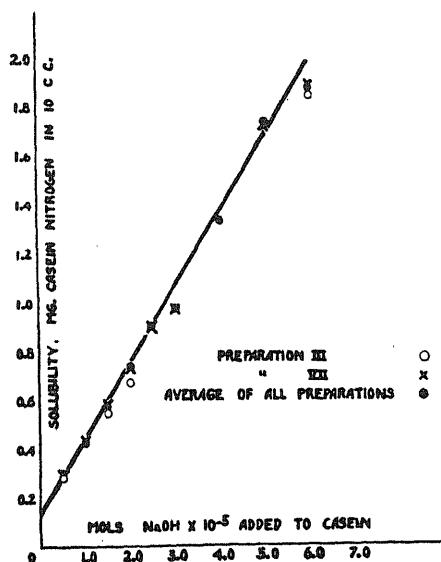


Figure 1. Solubility of casein. From Cohn, E. J., and Hendry, J. L., *J. Gen. Physiol.*, 5, 521 (1923).

to be approximately 2,100 grams of casein per mole of sodium hydroxide at 25°.

Casein has, as we have seen, a certain solubility in water at this temperature, estimated to be close to 18 mg nitrogen per liter (Table 1). Extrapolation with the straight line drawn in Fig. 1 indicates that casein to which no sodium hydroxide has been added has a solubility in water of 14 mg casein nitrogen per liter. The difference, 4 mg casein nitrogen, is considered a measure of the solubility of casein ions in equilibrium with the casein molecule. The solubility of the molecule,  $S_n$ , is considered to

out of serum proteins in phosphate, acetate, or citrate solutions and have differentiated several distinct steps in the process, presumably corresponding to the separation of different protein fractions. Recent studies of the salting out of the proteins of serum, separations being followed by electrophoretic and ultracentrifugal analyses<sup>49, 50, 51, 52, 53, 54</sup> indicate, however—as Hardy long ago suggested—that the effect of salts on serum proteins is very complex.\*

The underlying mechanism of the salting-out effect may be considered from varying points of view. Hofmeister<sup>58</sup> emphasized an important

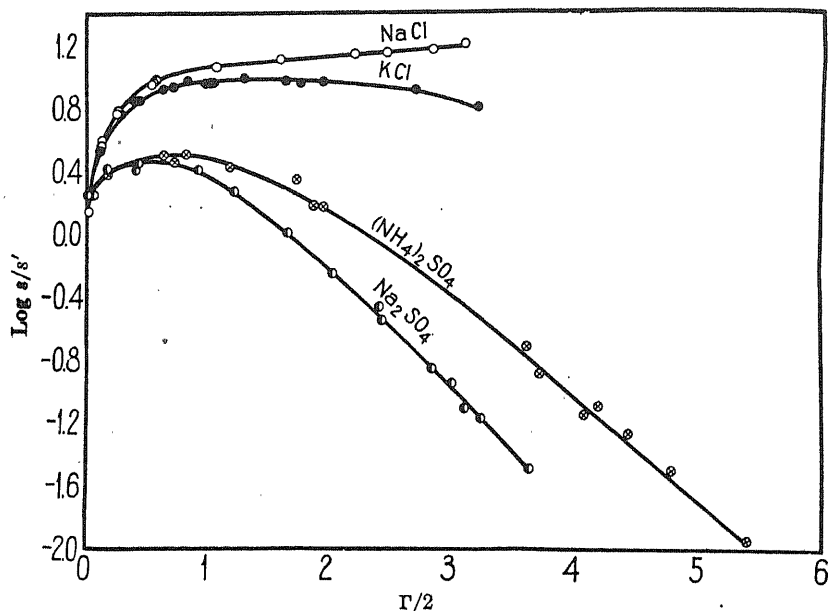


Figure 8. The solubility of hemoglobin in salt solutions of varying ionic strengths. Data from A. A. Green. From Cohn, E. J., *Chem. Rev.*, **19**, 241 (1936).

element of the truth in suggesting that the phenomenon was due to a "dehydration" of the protein by the added salt. This conception, couched in more explicit and quantitative terms, forms the basis of Debye's theory

<sup>49</sup> Mutzenbecher, P. v., *Biochem. Z.*, **266**, 226, 250, 259 (1933).

<sup>50</sup> McFarlane, A. S., *Biochem. J.*, **29**, 407, 660, 1175, 1209 (1935).

<sup>51</sup> Kekwick, R. A., *ibid.*, **32**, 552 (1938).

<sup>52</sup> Cohn, E. J., McMeekin, T. L., Oncley, J. L., Newell, J. M., and Hughes, W. L., *J. Am. Chem. Soc.*, **62**, 3386 (1940).

<sup>53</sup> McMeekin, T. L., *ibid.*, **62**, 3393 (1940).

<sup>54</sup> Cohn, E. J., Luetscher, J. A., Jr., Oncley, J. L., Armstrong, S. H., Jr., and Davis, B. D., *ibid.*, **62**, 3396 (1940).

\* A valuable discussion of the solubility of proteins is given in the monograph by Pauli and Valkó<sup>55</sup>. In this connection see also McMeekin, Luetscher, and Cohn<sup>56</sup>.

<sup>55</sup> Pauli, W., and Valkó, F., *Kristallisation der Eiweißkörper*, 2nd edition, Dresden and Leipzig (1933).

<sup>56</sup> McBain, J. W., and Jameson, E., *Trans. Faraday Soc.*, **26**, 768 (1930).

<sup>57</sup> Cohn, E. J., *Chem. Rev.*, **28**, 395 (1941).

<sup>58</sup> Hofmeister, F., *Arch. Exptl. Path. Pharm.*, **24**, 247 (1887-88).

of salting out<sup>59</sup>, according to which the salt ions attract around themselves the more polarizable molecules of the medium (in this case water), thereby squeezing out other components such as proteins. The specific influences of different salts, as enunciated by Hofmeister in his studies on proteins, are found also in the action of salts on amino acids, and indeed on the simplest gases<sup>60, 61</sup>.

The striking qualitative similarity in the action of salts on the solubility of cystine\* and of hemoglobin (Fig. 8) should be stressed. The salting-out effect in the action of salts on amino acids is apparent even in very dilute salt solutions; in proteins, likewise, its significance probably will prove to be even greater than earlier investigators were led to suspect. The elucidation of the nature of the effect is not a special problem, confronting the colloid chemist only; it is rather a problem of extraordinarily wide scope in the general theory of solutions.

#### VIII. SOLUBILITY IN DILUTE SALT SOLUTIONS: SALTING-IN

Neutral salts have a profound solvent action upon the class of proteins termed globulins. This effect, first noted by Denis, was investigated in 1905 by W. B. Hardy<sup>13</sup> and by Mellanby<sup>24</sup> in serum globulin, and by Osborne and Harris<sup>62</sup> for edestin. The changes in solubility observed are so large as to form the basis not only for classification but also for methods of purification. Presumably the dipole moments of these molecules are also very large, but they have not yet been investigated, nor have preparations of these proteins been available in such a state of purity that their solubility in solution was independent of the amount present in the solid phase.

Hemoglobin and lactoglobulin satisfy this criterion as closely as any proteins thus far investigated. Activity coefficients in aqueous sodium chloride of hemoglobin<sup>8, 33, 63</sup> and lactoglobulin<sup>64</sup> have been calculated from solubility measurements, and are graphically represented in Fig. 9, where they are compared with those of glycine, asparagine and cystine. As in the case of the peptides, the logarithm of the activity coefficient increases with the dipole moment, and as a first approximation, by slightly less than the first power of the dipole moment.

These results would lead one to conclude that globulins are proteins of high dipole moment. Their low solubility in water is presumably due, therefore, to very high crystal lattice energies for this class as for the

<sup>59</sup> Debye, P., and McAulay, J., *Physik. Z.*, 26, 22 (1925); Debye, P., *Z. physik. Chem.*, 130, 56 (1927).

<sup>60</sup> Scatchard, G., *Chem. Rev.*, 3, 333 (1927); *Trans. Faraday Soc.*, 23, 454 (1927).

<sup>61</sup> Randall, M., and Failey, C. F., *Chem. Rev.*, 4, 271 (1927).

\* Chapter II, Fig. 4.

<sup>62</sup> Osborne, T. B., and Harris, I. F., *Am. J. Physiol.*, 14, 151 (1905).

<sup>63</sup> Green, A. A., *J. Biol. Chem.*, 93, 495 (1931).

<sup>64</sup> Palmer, A. H., *J. Biol. Chem.*, 104, 359 (1934).



When heterogeneous equilibrium is established between the solvent and a single saturating body, the solubility,  $S$ , may be put equal to the sum of the solubility of the protein molecule,  $S_n$ , and of the concentrations of soluble protein anions,  $P^-$ ,  $P^{--}$ ,  $P^{---}$ ,  $P^{m-}$  and protein cations  $P^+$ ,  $P^{++}$ ,  $P^{+++}$ ,  $P^{n+}$ , where the positive or negative charges are considered net charges of protein ions present as soluble salts. We may therefore write:

$$S = S_n + (P^+ + P^-) + (P^{++} + P^{--}) + (P^{+++} + P^{---}) + \dots (P^{n+} + P^{m-}) \quad (1)$$

This equation does not specify whether the protein molecule be uncharged or a dipolar ion, or whether the ions have multiple charged groups and higher electrical moments, but merely that the molecule is neutral and the ions into which it dissociates vary in valence. The greatest valence of the cation is tentatively taken as the largest number of positively,  $n+$ , and of the anions of negatively,  $m-$ , charged groups of the protein.

Substituting for the concentration of the ions the mass law expressions for their concentration, we have, if we neglect activity coefficients:

$$\begin{aligned} S = S_n \left[ 1 + \left( \frac{(H^+)}{K_n} + \frac{K_{n+1}}{(H^+)} \right) + \left( \frac{(H^+)^2}{K_n K_{n-1}} + \frac{K_{n+1} K_{n+2}}{(H^+)^2} \right) \right. \\ \left. + \left( \frac{(H^+)^3}{K_n K_{n-1} K_{n-2}} + \frac{K_{n+1} K_{n+2} K_{n+3}}{(H^+)^3} \right) \right. \\ \left. + \dots \left( \frac{(H^+)^n}{K_n K_{n-1} K_{n-2} \dots K_1} + \frac{K_{n+1} K_{n+2} K_{n+3} \dots K_{m+n}}{(H^+)^m} \right) \right] \quad (2) \end{aligned}$$

Solubility of a protein even in the neighborhood of its isoelectric point is the sum of the solubility of the saturating body and of the ions into which it dissociates. The ratio of neutral molecules to ions depends upon the overlapping of the acid and basic dissociation constants. The summation of the first term in each parenthesis, the ratio of the hydrogen ion to the dissociation constant, measures the dissociation into cations; the summation of the second term in each parenthesis the dissociation into anions. If the hydrogen ion concentration is taken as that of the isoelectric point, the concentration of anions is equal to the concentration of cations, and the greater the constants defining the dissociation respectively into acids and bases the greater the number of ions in equilibrium in the isoelectric state with neutral molecules.

### III. SOLUBILITY AND THE HEAT OF NEUTRALIZATION

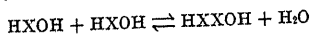
If base is added to such a protein salt, the first effect will generally be dissociation of carboxyl groups. The number of these losing protons—and rendering the protein a dipolar ion—will vary from protein to protein and depend both on the number of carboxyl groups and the strength of their dissociation. The heats of neutralization of most carboxyl groups is small and the number dissociating will thus not greatly vary with temperature.

As more base is added to the protein and the pH increases, the positively charged imidazole, amino and guanidino groups will tend to lose protons, reducing the number of protein cations and increasing the number of protein anions. As we have seen, the pH ranges in which the basic groups of peptides dissociate are far greater for guanidino than for amino, and for most amino than for imidazole groups. On the average, therefore, the first positive groups that lose protons, and which together with increase in the number of negatively charged groups, due to carboxyl dissociation, account for the decrease in positive and the increase in negative charge of proteins as the reaction becomes less acid and approaches neutrality, are the imidazole groups.

The heat of neutralization of imidazole groups is far greater than that of carboxyl groups (see Chapters 4 and 20). As a result the combining capacity of a protein will vary far more with temperature in neutral than in acid solution and more still in alkaline solutions in which the amino groups lose their charge. That this is true in homogeneous reactions has been demonstrated by EMF measurements (Chapter 20). Solubility measurements should also reflect influence of different temperature on the acid and basic groups of proteins. The nature of the effect to be expected on theoretical grounds is indicated by equation 2.

T. B. Robertson<sup>3</sup> early carried out an extensive series of measurements on the solubility of casein in solutions both of alkali and alkaline earth hydroxides\* at various temperatures. In the case of alkaline-earth

<sup>3</sup> Robertson, T. B., *J. Biol. Chem.*, 5, 147 (1908-9).  
 \*In 1888 Seldner<sup>1</sup> noted that the salts of casein with Ca, Ba, Sr, and Mg become opalescent when heated to about 49°. The opalescence disappeared upon the cooling of the solution. Osborne<sup>2</sup> confirmed these observations and also found that Li caseinate shows a very high solubility when warmed; ammonium caseinate, however, showed no sign of precipitation. Osborne explained this phenomenon in terms of hydrolysis of the protein salt: on heating the solution, casein is formed, which being insoluble, produces the opalescence. In 1908, in an extensive investigation upon the effect of temperature upon the solubility of casein with bases, Robertson<sup>3</sup> found that in the case of the bivalent bases, those containing this protein and monovalent bases, the solubility increases. Robertson<sup>3</sup> correctly pointed out that the hypothesis of the rise in temperature the solubility increases. Osborne did not explain the effect of temperature upon the solubility of casein in monovalent bases" p. 368-369 and concluded that the effect of temperature upon the solubility of casein in monovalent bases can be explained on the supposition that the effect of temperature consists in:  $HXOH + HXOH \rightleftharpoons HXXOH + H_2O$



towards the right so that a given amount of alkali salt is associated with a molecule of nearly double the weight, neutralizes nearly twice as much acid at 21° as it does at room temperature (21°). The marked distinction in the solubility of casein in  $Ca(OH)_2$  solutions, which occurs on raising the temperature, can be explained on the supposition that the salt  $Ca(HXXOH)_2$  is soluble." p. 369



hydroxides, unlike sodium hydroxide, there is evidence that new saturating bodies are formed and the conditions under the phase rule for two saturating

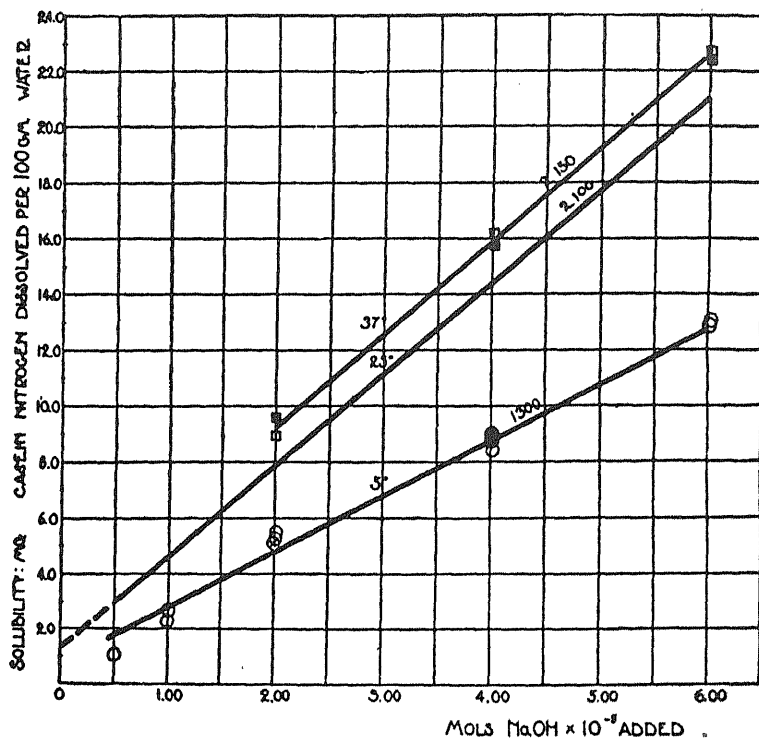


Figure 2. The solubility of casein at 5°, 25°, and 37° in small amounts of NaOH. From Pertzoff, V., *J. Gen. Physiol.*, 10, 961 (1926-27).

TABLE 4. THE EQUIVALENT WEIGHT OF CASEIN FROM SOLUBILITY MEASUREMENTS AT VARIOUS TEMPERATURES

Investigator	Temperature (°C.)	Equivalent weight (gm.)	Remarks
Author.....	5.0	1300	With small amounts of base
Robertson (1).....	24.0	2000	
Cohn and Hendry (2).....	25.0	2100	Average: 2070
Author.....	37.0	2150	
Author.....	49.0	2600	
Robertson (1).....	66.0	3700	From 60° to 85° the equivalent weight again remains practically constant

From Pertzoff, V., *J. Gen. Physiol.*, 10, 961 (1927)

(1) Robertson, T. B., *J. Biol. Chem.*, 5, 147 (1908-09).  
 (2) Cohn, E. F. and Hendry, V. *J. Gen. Physiol.*, 5, 501 (1922-23).

spherical coordinates. This resembles more closely a globular protein molecule, which possesses a considerable number of charges, determined by the number and locations of ionized groups on acidic and basic amino acid residues.

The mutual electrostatic energy of such a complex dipolar ion and electrolyte ions in a dielectric medium was evaluated, and, assuming that the only deviation from ideality in the chemical potential arises from this electrostatic interaction, the activity coefficient of the complex ion was derived. The result is as follows<sup>(67, see equation 21)</sup>:

$$\ln \gamma = -\frac{Q_0}{2DkT} \cdot \frac{\kappa}{1 + \kappa a} - \frac{\kappa^2}{2DkT} \sum_{n=1}^{\infty} \left( \frac{(2n+1)Q_n}{(2n-1)(n+1)^2 a^{2n-1}} \times \frac{K_{n-1}(\kappa a)}{K_{n+1}(\kappa a) + \frac{nb^{2n+1}\kappa^2 K_{n-1}(\kappa a)}{(n+1)(2n-1)(2n+1)a^{2n-1}}} \right) \quad (10)$$

Here  $\gamma$  is the activity coefficient of the complex ion;  $D$  is the dielectric constant of the solvent;  $k$  the Boltzmann constant;  $T$  the absolute temperature;  $\kappa$  is the Debye-Hückel parameter  $[(4\pi N e^2 / 1000 D k T) \sum_i z_i^2 C_i]^{1/2}$ , where  $N$  is Avogadro's number,  $e$  the electronic charge,  $z_i$  the valence of the  $i$ 'th ionic species, and  $C_i$  its concentration in moles per liter. The parameter  $\kappa$  is accordingly proportional to the square root of the ionic strength. Continuing with definitions,  $b$  is the radius of the sphere, and  $a$  the mean distance of closest approach of the salt ions, *i.e.*,  $b$  plus the mean ionic radius. The  $Q_n$ 's are the double sums

$$Q_n = \sum_{k=1}^M \sum_{l=1}^M e_k e_l r_k^n r_l^n P_n(\cos \theta_{kl}) \quad (11)$$

where  $M$  is the total number of charges imbedded in the spherical complex ion, and  $\theta_{kl}$  is the angle subtended at the center of the sphere by the  $k$ 'th and  $l$ 'th charges, which have charges  $e_k$  and  $e_l$  respectively, and are at distances  $r_k$  and  $r_l$  from the center of the sphere. The  $P_n$ 's are the Legendre functions. Specifically,  $Q_0$  is the square of the net charge of the complex ion, and  $Q_1$  is the square of the dipole moment. The succeeding  $Q_n$ 's involve the higher order multipole moments.\* Finally, the functions  $K_n$  are polynomials given by

$$K_n(x) = \sum_{s=0}^n \frac{2^n n! (2n-s)!}{s! (2n)! (n-s)!} x^s \quad (12)$$

\* When the charges are all taken as integral multiples of the electronic charge,  $e(e_i/e = z_i)$ , imbedded at the same distance from the center,  $r$ , equation 11 can be conveniently rewritten

$$Q_n = e^2 r^{2n} \sum_k \sum_l z_k z_l P_n(\cos \theta_{kl})$$

Here the double sum is a pure number which depends only upon the geometrical relations in the charge configuration.

combining weight of casein and sodium hydroxide, in excellent agreement with that considered above and with results of Pertzoff<sup>1</sup>. These are graphically represented in Fig. 2 and illustrate the contribution respectively of neutral molecules and of ions to solubility. The parallel straight lines, drawn through the measurements at 25° and 37° indicate a difference in the solubility of the neutral molecule, but not in the equivalent combining weight. The measurements at 5° suggest both a lower solubility for the neutral molecule and a lower combining weight. The equivalent combining weights estimated from solubility measurements at various temperatures are given in Table 4. The measurements at higher temperatures can less readily be compared with those below 40°, since changes in the casein molecule may well take place at the higher temperatures.

#### IV. SOLUBILITY PRODUCT CONSTANT

At reactions sufficiently removed from the isoelectric point some of the terms in equation 2 might be expected to be negligible. Thus at sufficiently acid reactions the second term in each parenthesis, giving the concentration of protein anions, vanishes, whereas at reactions appreciably alkaline to the isoelectric point the first term might be expected to vanish. Moreover, the higher terms might be expected to be less important than those describing groups dissociating with constants close to the hydrogen ion concentration of the neutral protein.

The tetrapole cystine may be considered a prototype for protein behavior in this, as in other respects. Its solubility has been studied as a function of pH by Sano<sup>6</sup> and is satisfactorily described by equation 2 with the terms containing the first and second dissociation constants respectively as acids and bases included.

The studies upon the solubility of casein in systems containing bases (Table 3) were described, however, by the terms for dissociation as a divalent acid.

$$S = S_n \left[ 1 + \left( \frac{K_{a1}K_{a2}}{(H^+)^2} \right) \right] \quad (3)$$

The solubility measurements reported in Table 3 have been analyzed in terms of the above equation and values of the product  $S_n K_{a1} K_{a2}$  estimated to be  $2.2 \times 10^{-12}$  gram casein per liter at 25°. The solubility of the molecule,  $S_n$ , was estimated to be 0.09 gram per liter, and the product of the dissociation constants,  $K_{a1} K_{a2}$ , to be  $24 \times 10^{-12}$ .

The solubility of casein in HCl has been studied by Linderstrøm-Lang and Kodama<sup>7</sup>. Although they found the solubility of their casein preparations in these solvents to depend upon the amount of saturating body

Equation 10, then, evaluates the activity coefficient, in salt solution, of a spherical molecule containing any given number of charges arranged in any given pattern. When the net charge is zero, as in an isoelectric dipolar ion, the term  $Q_0$  vanishes; and the logarithm of the activity coefficient is linear in the ionic strength in very dilute salt solutions, although it departs from linearity at higher concentrations.

Were the configuration of charges on the surface of an isoelectric protein molecule known and mapped out, it would be possible to predict the salting in by Kirkwood's equation, insofar as the molecule could be regarded as possessing a spherical shape (and in certain cases ellipsoidal shapes; see Chapter 12). Unfortunately the reverse problem is indeterminate. Given an experimental salting-in function, there are many different possible configurations of charges which could be responsible for it, and it is impossible to decide among them.

It is therefore necessary to seek independent experimental evidence concerning the configuration of charges. Any configuration can be described by a series of terms giving its dipole moment, quadrupole moment, and so on. Experimental methods can thus far furnish only the first of these terms—the dipole moment, which is derived from dielectric constant measurements. The simplest possible charge configuration corresponding to an experimental dipole moment is a pair of charges of appropriate magnitude, and opposite sign, placed at opposite poles of the spherical molecule. This has been the model taken for calculations of the salting in of hemoglobin and lactoglobulin. Of course there are many other possible configurations which could give the same dipole moment; but these could be identified only by their higher moments.

### Hemoglobin

The experimental measurements by Ferry, Cohn and Newman<sup>69</sup> of the solubility of horse carboxyhemoglobin in ethanol-water mixtures containing sodium chloride give the salting-in functions shown in Fig. 10, curves 2 and 3. Here  $-\log \gamma$  is multiplied by  $D/D_0$  and  $\Gamma/2$  by  $D_0/D$ , where  $D_0$  is the dielectric constant of water, and  $D$  that of the particular ethanol-water mixture employed (25% and 35% ethanol respectively). This procedure serves to reduce both dielectric constants to that of water, inasmuch as the dielectric constant enters in Kirkwood's equation as a factor of  $1/D^2$ . The failure of this reduction to make curves 2 and 3 coincide is probably to be explained by the salting-out correction, as discussed below.

For a theoretical calculation of the salting in of hemoglobin, the molecule may be considered to be a sphere. The experimental value of the dipole moment is 500 Debye units<sup>70</sup>. We take two pairs of elementary charges

<sup>69</sup> Ferry, R. M., Cohn, E. J., and Newman, E., *J. Biol. Chem.*, **114**, xxxiv (1936).

<sup>70</sup> Oncley, J. L., *J. Am. Chem. Soc.*, **60**, 1115 (1938).

pretable on the assumption that "on the acid side of its isoelectric point casein also behaves as though it were divalent"<sup>8</sup> (p. 538). Solubility on both sides of the isoelectric point is given by an equation of the form:

$$S = S_n \left[ 1 + \frac{(H^+)^2}{K_1 K_2} + \frac{K_3 K_4}{(H^+)^2} \right] \quad (4)$$

Solubility measurements in acid solution (where the total protein in each system was 8.2 grams per liter) plotted as ordinate against the square of the hydrogen ion fall on a straight line (Fig. 3), and this was true also

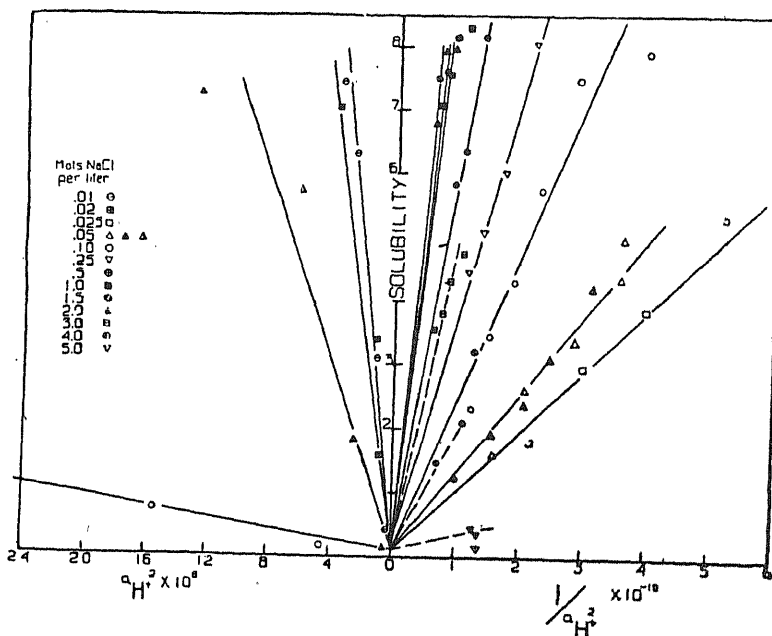


Figure 3. The solubility of casein at varying pH from the experiments of Linderström-Lang and Kodama and of Sørensen and Sládek. Taken from Green, A. A., *J. Biol. Chem.*, **93**, 517 (1931).

when the systems contained concentrations of NaCl. The comparable studies of Sørensen and Sládek<sup>9</sup> on the solubility of casein in NaCl and NaOH (where the total protein in each system was close to 8.5 grams per liter) are also plotted in Fig. 3 against  $(1/(H^+)^2)$ .

#### V. SOLUBILITY AND THE ACTIVITY COEFFICIENTS OF PROTEIN IONS

Equation 2 was developed from equation 1 on the assumption that the activity coefficients of protein ions and dipolar ions could be set equal to



placed at opposite poles and imbedded to the same extent—estimated to be 1.2 Å from the surface as in glycine. This gives a distance of separation of 51.6 Å, and the charges of  $+2e$  and  $-2e$  at this distance produce a dipole moment of 495 Debye units, which corresponds closely to the experimental value. For this type of configuration the even terms in Kirkwood's equation 10 vanish. Summing the odd terms through the seventh, we obtain the theoretical curve 1, Fig. 10. This lies close to the experimental curves.

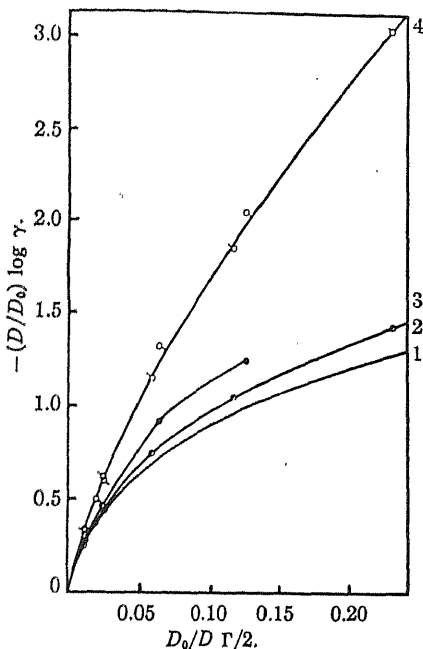


Figure 10. Activity coefficients of carboxyhemoglobin in sodium chloride at  $-5^{\circ}$ : 1, calculated from Kirkwood's equation on the assumption of a molecular weight of 66,700 and a dipole moment of  $500 \times 10^{-18}$  esu.; 2,  $\odot$ , observed in 25% ethanol; and 3,  $\bullet$ , in 35% ethanol; 4, measurements in 25% ethanol;  $\circ$ , and in 35% ethanol;  $\alpha$ , corrected by a "salting out" constant of 7.9. From Ferry, R. M., Cohn, E. J., and Newman, E. S., *J. Am. Chem. Soc.*, **60**, 1480 (1938).

Provided the latter are corrected for salting out, however, the agreement is less satisfactory.

The salt interacts with the protein not only to increase solubility in the manner described by equation 10, but also to decrease solubility because of cavities in the solvent represented by the molecules of the protein. The decrease in solubility, or salting out, is given by

$$(D/D_0) \log \gamma = K_s \Gamma/2 \quad (13)$$

where  $K_s$  is calculated from another equation of Kirkwood<sup>71</sup>.

$$K_s = \frac{4\pi N\epsilon^2}{2303D_0 kT} \frac{b^3}{a} \frac{D-1}{2D+1} \quad (14)$$

Here only the size of the sphere is involved, and not its charge configuration.

The figure for  $K_s$  calculated from this equation for hemoglobin is 7.90. When values of  $K_s$   $\Gamma/2$  are added to the experimental values of  $-(D/D_0) \log \gamma$  in Fig. 10 the results for hemoglobin in 25% and 35% ethanol nearly coincide (curve 4). They are, however, far above the theoretical curve 1. The discrepancy may be attributed to neglect of higher moments in the theoretical calculations, because of the simplified charge configuration assumed, or to ion-ion interactions.

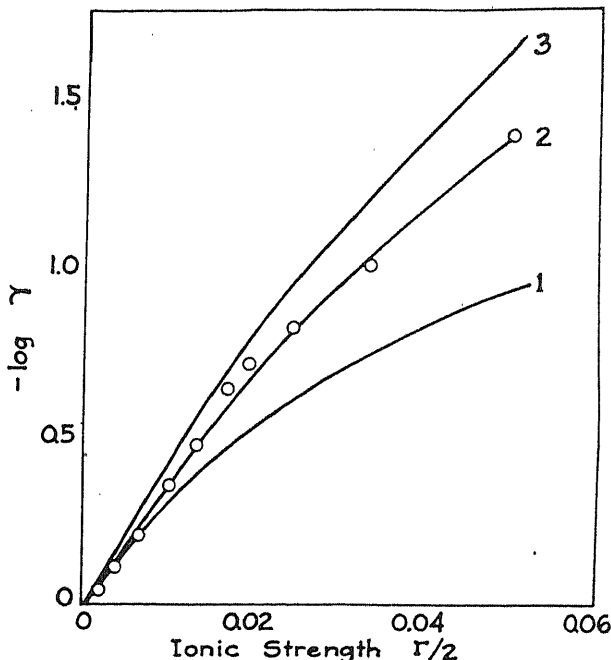


Figure 11. Activity coefficients of lactoglobulin in sodium chloride solutions at 30°

### Lactoglobulin

The measurements of Palmer<sup>64</sup> give for the solubility of lactoglobulin in dilute sodium chloride solutions at 30° the experimental curve, 2, given in Fig. 11 (cf. also Fig. 9), which is far steeper than for hemoglobin.

For a theoretical calculation, the same model as for hemoglobin has been assumed except for the magnitude of the charges<sup>71a</sup>. The experimental

<sup>71</sup> Kirkwood, J. G., Personal communication. See also Chapter 12.

<sup>71a</sup> We are indebted to M. Jean Mayer for carrying out these calculations.

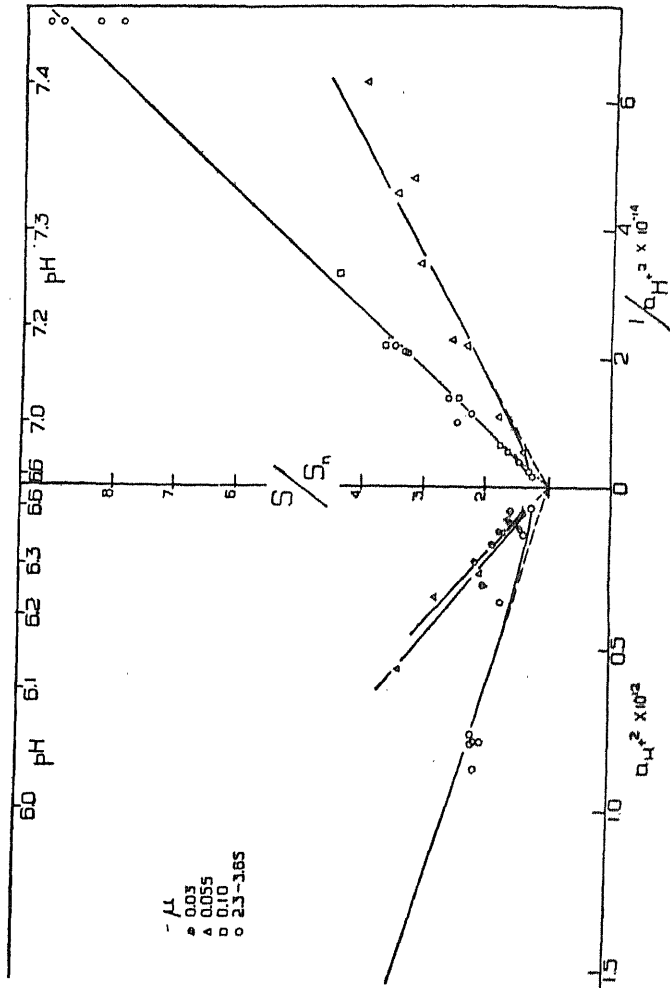
unity. This assumption is clearly not valid for any appreciable ionic strength, concentration of neutral molecules or of dipolar ions, change in temperature or in dielectric constant from that of the standard state. Introducing activity coefficients in the mass law expressions for the concentration of the ions we have, instead of equation 2, the following:

$$\begin{aligned}
 S = S_n \left( 1 + \gamma_n \left[ \left( \frac{(H^+)}{K_n \gamma_{P^+}} + \frac{K_{n+1}}{(H^+) \gamma_{P^-}} \right) + \left( \frac{(H^+)^2}{K_n K_{n-1} \gamma_{P^{++}}} + \frac{K_{n+1} K_{n+2}}{(H^+)^2 \gamma_{P^{--}}} \right) \right. \right. \\
 + \left( \frac{(H^+)^3}{K_n K_{n-1} K_{n-2} \gamma_{P^{+++}}} + \frac{K_{n+1} K_{n+2} K_{n+3}}{(H^+)^3 \gamma_{P^{---}}} \right) \\
 \left. \left. + \dots \left( \frac{(H^+)^n}{K_n K_{n-1} K_{n-2} \dots K_1 \gamma_{P^{n+}}} + \frac{K_{n+1} K_{n+2} K_{n+3} \dots K_{m+n}}{(H^+)^n \gamma_{P^{m-}}} \right) \right] \right) \quad (5)
 \end{aligned}$$

The form of this equation is entirely comparable to that of equation 2, but the terms within the parentheses, giving the concentrations of proteins as monovalent anions and cations, divalent, trivalent and higher valent anions and cations, are multiplied in each case by the activity coefficient of the neutral molecule or dipolar ion,  $\gamma_n$ . The only other change from equation 2 is that instead of thermodynamic dissociation constants we have apparent dissociation constants; that is, dissociation constants multiplied by activity coefficients which define deviation from behavior in the standard state, and are generally written as  $K'$  (see Chapters 3 and 4). Solubility measurements, like EMF measurements, can thus yield thermodynamic dissociation constants only by extrapolation to conditions such that the activity coefficients may be taken as unity.

The most complete study of the solubility of a protein, as a function of pH, ionic strength and temperature, is that of Green on the carboxy-hemoglobin of the horse<sup>8</sup>. The solubility of the carboxyhemoglobin was, within the limits of error of the measurements, independent of the amount of saturating body in the system. The hydrogen ion was very accurately controlled by the use of phosphate buffers of known pH and ionic strength. The solubility,  $S$ , at 25°, of the hemoglobin in water was estimated to be 17 grams per liter, and of the neutral molecule,  $S_n$ , 11 grams per liter at zero ionic strength. When the extensive solubility measurements reported for solutions more acid than the isoelectric point were plotted against various powers of the hydrogen ion activity or for solutions more alkaline than the isoelectric point against various powers of the reciprocal of the hydrogen ion concentration, it was found that the measurements fell on straight lines if all terms other than those containing the second power of the hydrogen ion activity were neglected. These studies and those

Figure 4. The solubility of hemoglobin in phosphate buffers of constant ionic strength and varying pH. From Green, A. A., *J. Biol. Chem.*, **93**, 517 (1931). The symbol  $\mu$  denotes ionic strength.



dipole moment is 730 Debye units at  $25^{\circ}$ <sup>72</sup>; three pairs of charges at a distance of 43 Å give a moment of 618 D, which is the nearest approximation that can be made with an integral number of elementary charges. Summing the first, third, and fifth terms in Kirkwood's equation 10, we obtain the theoretical curve, 1, Fig. 11. This is considerably higher than the corresponding curve for hemoglobin, because of the smaller size of the molecule, and the greater dipole moment. As in the case of hemoglobin, the theoretical curve lies close to the experimental, especially at low ionic strengths, until the salting-out correction is applied.

The experimental curve can now be corrected for salting out, as was done in the study of hemoglobin. The value of  $K_s$  calculated from equation 14 is 5.61. When values of  $K_s\Gamma/2$  are added to the experimental values of  $-\log \gamma$ , curve 3 is obtained. Here the salting-out correction makes much less contribution than for hemoglobin, owing to the smaller

TABLE 7. NEGATIVE LOGARITHMS OF ACTIVITY COEFFICIENTS CALCULATED FROM KIRKWOOD'S EQUATION\* FOR SPHERICAL MOLECULES OF MOLECULAR WEIGHT 40,000

Ionic Strength $\Gamma/2$	1 Pair Charges $\mu = 206D$	2 Pairs Charges $\mu = 412D$	3 Pairs Charges $\mu = 618D$	4 Pairs Charges $\mu = 824D$	5 Pairs Charges $\mu = 1030D$
		$-\log \gamma$ : Calculated			
0.001	0.0043	0.0171	0.0384	0.0681	0.107
0.01	0.0337	0.135	0.304	0.538	0.842
0.05	0.103	0.411	0.924	1.64	2.57
0.1	0.163	0.65	1.47	2.61	4.08
0.2	0.227	0.91	2.05	3.64	5.69

\* Same as equation 10 of this chapter.

size of the molecule. The corrected curve, 3, is somewhat higher than the theoretical salting in, but the difference is less than for hemoglobin.

### Influence of Dipole Moment on Salting-In

As a general survey of the effect of dipole moment upon the salting in of a protein, as predicted by Kirkwood's equation 10, calculations have been made<sup>71a</sup> for spherical molecules of molecular weight 40,000 and various dipole moments, at  $25^{\circ}$ . The configuration taken, as before, was a pair of charges at opposite poles and imbedded at 1.2 Å below the surface. Their magnitudes range from 1 to 5 electronic charges. The case of 3 charges is, of course, that of lactoglobulin discussed above. The odd terms of the equation through the fifth were summed. The results are given in Table 7 and are plotted in Fig. 12.

The values of  $-\log \gamma$  at a given salt concentration increase with the square of the dipole moment, for this simple charge configuration. For the highest moment taken, 1030 Debye units, the salting in is enormous,

<sup>72</sup> Ferry, J. D. and Oncley, J. L., *J. Am. Chem. Soc.*, **63**, 272 (1941).

equation 4, but containing terms for the activity coefficients both of the ions and the dipolar ions.

$$S = S_n \left[ 1 + \gamma_n \left( \frac{(H^+)^2}{K_1 K_2 \gamma_{P^{++}}} + \frac{K_3 K_4}{(H^+)^2 \gamma_{P^{--}}} \right) \right] \quad (6)$$

Therefore the slope in acid solution  $[(S/S_n) - 1]/(H^+)^2 = \gamma_n(1/K_1 K_2 \gamma_{P^{++}})$ , and in alkaline solution  $[(S/S_n) - 1]/(H^+)^2 = \gamma_n(K_3 K_4/\gamma_{P^{--}})$ , where  $\gamma_n S_n = S_n^0$ , the solubility of the protein molecule at zero ionic strength.

The use of this simplified expression instead of equation 5 does not explain, however, why the more complete expression is not necessary for the description of the behavior of these proteins in solutions varying in pH

TABLE 5. APPARENT DISSOCIATION CONSTANTS OF CARBOXYHEMOGLOBIN ESTIMATED FROM SOLUBILITY MEASUREMENTS IN PHOSPHATE BUFFER SOLUTIONS VARYING IN pH AND IONIC STRENGTH\*

Ionic Strength	Solubility of Neutral Molecules estimated from Equilibrium with		Dissociation Constants		Apparent Isoelectric Point $\frac{pK_1'K_2' + pK_3'K_4'}{4}$
	Cations $S_n$	Anions $S_n$	Cations $pK_1'K_2'$	Anions $pK_3'K_4'$	
0.014-0.015	17.9		12.68		
0.026-0.031	19.8		12.64		
0.039-0.065	21.7	21.4	12.69	14.25	6.74
0.082-0.103		24.3		13.96	
0.160	29.6		12.33		
0.314-0.332	36.3	35.9	12.24	14.10	6.59
0.645	44.5		11.95		
1.49	20.3		12.10		
1.56		18.2		14.07	
2.34-3.65			12.20	13.96	6.54

\* Green, A. A., *J. Biol. Chem.*, **93**, 517 (1931).

and ionic strength. The graphical representation of the hemoglobin solubility measurements in Fig. 4 may be compared with those upon casein in Fig. 3. Since the ordinates in Fig. 4 are taken as  $S/S_n$ , the slopes of the straight lines at the right of the figure yield the ratio of the activity coefficient of the neutral molecule to that of the protein anion multiplied by the dissociation constants of the latter, whereas those on the left yield the ratio of the activity coefficients of the neutral molecule to the product of the protein cation and its dissociation constants. Analyzed in these terms these results yield the constants in the following table (Table 5).

Equation 4 thus appears applicable to casein and hemoglobin in the presence of salt, provided the values of  $S_n$  and  $K$  are not regarded as con-

representing nearly a ten-fold increase in solubility at an ionic strength of 0.01. It will be of interest to compare these figures with experimental measurements on various proteins of molecular weight in the neighborhood of 40,000, including egg albumin, zein, and insulin, as well as lactoglobulin, when both solubility and dipole moment data are available.

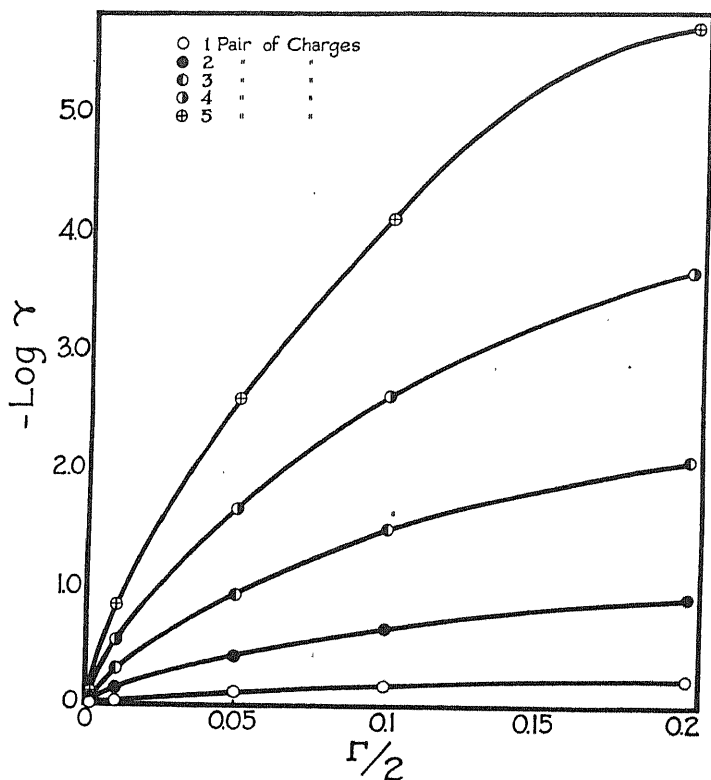


Figure 12. Influence of dipole moment on activity coefficient

### Influence of Protein Ions on Salting-In

The observed values of  $-\log \gamma$  must represent not only interactions between salt ions and protein dipoles, but between salt ions and protein ions that are in equilibrium with the isoelectric protein. The number of such ions, and their valence, will increase, as we have seen, with distance from the isoelectric point, and with steepness of the titration curve. The nature of such interactions has been considered by Linderstrøm-Lang<sup>73</sup>. They cannot, of course, be neglected in interpreting the electrostatic forces between salts and proteins in the neighborhood of protein isoelectric points.

<sup>73</sup> Linderstrøm-Lang, K.; see Chapter 20, pp. 462, 468-77

hemoglobin that  $S_n$  at first increases on addition of salt, then passes through a maximum and, in concentrated solutions, decreases because of the salting-out effect. This corresponds to the characteristic solubility curve for a globulin in solutions of varying ionic strength. The coefficients  $K_1'K_2'$  and  $K_3'K_4'$  increase on addition of salt, slightly in hemoglobin solutions, very markedly in casein solutions. This increase of the constants with salt concentrations has a twofold consequence: (a) the value of the hydrogen ion activity for minimum solubility is given, for systems in which equations 4 and 6 hold, by the relation:

$$(H^+)^4 = K_1'K_2'K_3'K_4' \quad (7)^*$$

Since all of the  $K$ 's increase with increasing ionic strength, the activity of the hydrogen ion,  $(H^+)$ , at minimum solubility must increase also; or the pH of minimum solubility decreases. This effect is comparatively slight for hemoglobin; but for casein, Green has calculated from the data of Sørensen and of Linderstrøm-Lang that the pH of minimum solubility shifts from 4.8 in the absence of salt to 4.06 at ionic strength 0.1†. (b) Solubility, at constant pH, on the alkaline side of the isoelectric point is markedly increased by salt at low or moderate salt concentrations (since both  $\gamma_n$  and  $\gamma_{P--}$  in equation 6 decrease with ionic strength). This effect is found in both casein and hemoglobin. On the acid side of the isoelectric point,  $\gamma_n$  decreases,  $\gamma_{P++}$  increases, with increasing ionic strength. The former tends to increase the solubility,  $S$ , the latter to decrease it. In hemoglobin, at constant pH, on the acid side of the isoelectric point, the resultant effect is a very slight increase in solubility with increase in salt concentration, followed by salting out which begins at a relatively low salt concentration. In casein the increase of  $\gamma_{P++}$  is very rapid, and the casein cation is salted out very rapidly even by the smallest concentrations of salt, in contrast to the anion which is readily dissolved by salt under the same conditions.

Phenomena of this type have been clearly perceived by earlier observers of other proteins. Thus Osborne<sup>12</sup> wrote that "edestin dissolved in the least possible quantity of hydrochloric acid necessary for its solution is precipitated by traces of most mineral salts, but a slight excess of acid requires the addition of more salt for precipitation. The precipitation of

\* At the point where  $S$  is a minimum,  $dS/d(H^+)$  must be zero, and from equation 6:

$$\frac{dS}{d(H^+)} = S_n \left( \frac{2(H^+)}{K_1'K_2'} - \frac{2K_3'K_4'}{(H^+)^3} \right) = 0 \quad (8)$$

from which equation 7 follows. It is clear from the nature of equation (8) that equation 8 gives the conditions for a minimum and not for a maximum.

† Whether the isoelectric point as determined from cataphoresis measurements shifts similarly is a problem considered in the next chapter. It should be noted that the isoelectric point has been drawn between the "isoelectric" and "isoelectric" points in the diagram of Sørensen and Linderstrøm-Lang, and



Ion-dipole interaction cannot be considered constant over any appreciable pH range, for the ionization of different residue groups of the protein on either side of the isoelectric point can result either in increase or decrease in the dipole moment of the protein. Such changes in the dipole moment of the protein should also be reflected in dielectric constant measurements and such measurements for egg albumin<sup>74</sup> and  $\gamma$ -pseudoglobulin<sup>75</sup> have been reported. Only the simultaneous measure of the activity coefficient of the protein, its combination with acids and bases and the resulting change in dielectric constant can permit the estimation of the role of ion-ion and ion-dipole forces in determining the solubility and other relations of proteins and salts.

#### IX. SOLUBILITIES IN SOLUTIONS CONTAINING MORE THAN ONE DIPOLAR ION

Not only ions but other dipolar ions affect the solubility of proteins. The influence of other amino acids on the solubility of cystine is discussed in Chapter 11. Amino acids also influence the solubility of proteins and this effect is illustrated, for the few cases that have thus far been investigated, in Fig. 9 and compared with the effects of salts on the same proteins.

#### Hemoglobin

A beautiful study of the influence of glycine on the solubility of hemoglobin has been reported by M. M. Richards<sup>76</sup>. The logarithms of the solubility ratios of hemoglobin reported by this investigator are plotted in Fig. 9b against the glycine concentrations as abscissa, without correcting for the influence of the hemoglobin upon the dielectric constant of the solution. The solubility of hemoglobin in the absence of glycine in these experiments varied from 13.9 to 17.8 gm per 1000 grams of water, and was 60 gm per 1000 grams of water in 1.5 molal glycine. Since the dielectric increment<sup>70</sup> per gram of hemoglobin may be taken as 0.33, the dielectric constants of the saturated solutions were appreciably influenced by the protein as well as by the glycine. This effect had previously been considered with respect to the studies of Green<sup>33</sup> on the interaction of hemoglobin with sodium chloride in aqueous solution<sup>77</sup>. Correcting for it would yield still greater estimates of activity coefficients due to Coulomb forces.

The curves relating  $-\log \gamma$  and glycine concentration are very similar for the dipolar ions glycine and asparagine, for the tetrapolar ion cystine, and for the multipolar protein hemoglobin. In these studies the greater the interaction the greater the moments of the molecules. Thus the dipole moment of glycine and asparagine may be taken as 15, that of cystine as

<sup>74</sup> Dunning, W. J., and Shutt, W. J., *Trans. Faraday Soc.*, **34**, 479 (1938).

<sup>75</sup> Oncley, J. L., Bateman, J. B., Feeley, C., and McAn, M., Unpublished measurements.

<sup>76</sup> Richards, M. M., *J. Biol. Chem.*, **122**, 527 (1938).

<sup>77</sup> Cohn, E. J., *Chem. Rev.*, **19**, 241 (1936).

that he studied "seemed to vary with the buffer used". It thus becomes necessary to differentiate between the influence of neutral salts on the flocculation of edestin and on the insoluble compounds that this, and presumably other globulins, form with acids. It would seem as though Osborne's analytical procedure was less equivocal than the methods that have occasionally been applied to globulins without adequate consideration of their special chemical characteristics.

Edestin is possessed of but a small number of acid and basic groups that dissociate in the neighborhood of its isoelectric point. Osborne's measurements, interpreted on the basis of a molecular weight of 336,000, indicate that edestin contains 24 groups which can combine base in the neighborhood of the isoelectric point to form a soluble compound, and 24 which combine with acid to form the insoluble compound referred to above. A combination of acid with 24 other groups transforms the insoluble into a soluble hydrochloride<sup>14a</sup>. In all, therefore, 72 groups dissociate from pH 4.5 to 8.5.

Not only globulins but also albumins form insoluble salts with anions. Thus serum albumin is isoelectric near pH 4.8. It has generally been crystallized near this reaction from concentrated salt solutions; minimum solubility shifting somewhat to more acid reactions, the greater the ionic strength. Recently the observation was made by J. D. Ferry<sup>18</sup> that a fraction of serum albumin can also be crystallized from salt-free solutions near pH 4 as a somewhat insoluble sulfate—more soluble, however, at higher than at lower temperatures—and this phenomenon has been employed in the further fractionation and purification of the crystalline serum albumins<sup>19</sup>.

Specific interactions between protein cations and other anions, especially complex anions, have, of course, always been recognized and many of these lead to the formation of new saturating bodies. Many analytical and industrial processes, among them the combination of proteins with tannates, picrates, picrolonates and phosphotungstates, depend upon the formation of such insoluble protein salts. Combination of proteins with such anions often leads to changes in the proteins, however, which are not readily reversible.

A great many proteins are insoluble over a very wide pH range throughout which they combine with acids and bases to form a series of insoluble salts. The interpretation of the behavior of such protein solid phases demands a knowledge not only of the neutral molecule, but of the stoichiometric and electrochemical valences of the several insoluble compounds.

<sup>14a</sup> J. D. Ferry, *Trans. Faraday Soc.*, **36**, 100 (1940).

<sup>18</sup> J. D. Ferry, *Trans. Faraday Soc.*, **36**, 100 (1940).

<sup>19</sup> J. D. Ferry, *Trans. Faraday Soc.*, **36**, 100 (1940).

27, and that of hemoglobin as 500 Debye units. These interactions are, as a first approximation, proportional to the first power of the moment, whereas theoretical considerations have suggested<sup>78, 79</sup> that they should be proportional to a higher power and diminish with increase in volume of the molecule on the basis of models which, it is true, cannot be expected satisfactorily to reproduce all the characteristics of dipolar ions.

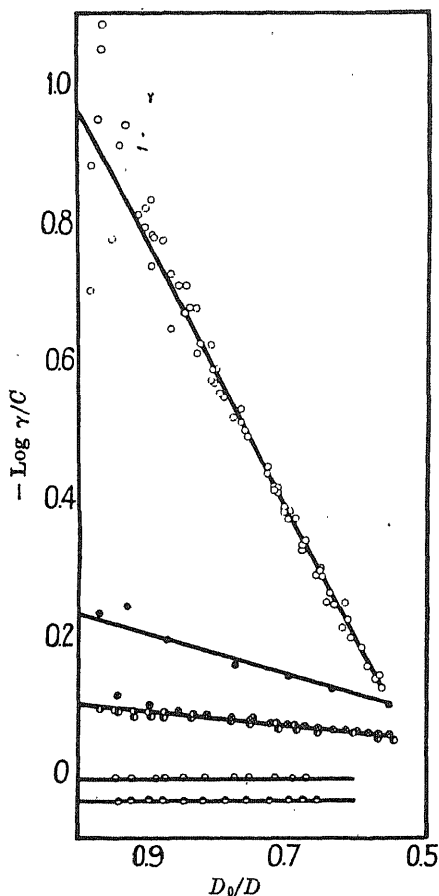


Figure 13. Activity coefficient in solutions of dipolar ions. The straight line drawn for each substance represents the equation:  $-(\log \gamma)/C = K_B^*(D_0/D) - K_S^*$ .

- , Hemoglobin;
- , Cystine;
- ⊕, Glycine;
- ⊙, Glycine;
- ⊗, Glycine;
- ⊖, Alanine;
- ⊙, α-Aminobutyric acid

From Cohn, E. J., McMeekin, T. L., Ferry, J. D., and Blanchard, M. H. *J. Physical Chem.* **43**, 182 (1939)

The interactions between glycine and other amino acids and peptides, whose structures are, to be sure, far better known than are those of the proteins, are considered in Chapter 10. Table 9 of that chapter also includes constants for the interaction of glycine and hemoglobin.

<sup>78</sup> Fuoss, R. M., *J. Am. Chem. Soc.*, **58**, 982 (1936).

<sup>79</sup> Kirkwood, J. G., *Chem. Rev.*, **19**, 275 (1936). See Chapter 12.

### Lactoglobulin

The influence of glycine upon the solubility of lactoglobulin<sup>80</sup> has been determined at 5° and 25°. At a concentration of 0.5*M* glycine, the solubilities of lactoglobulin at these temperatures are 6.8 and 9.1 g per liter, respectively, in water. The limiting slopes of  $-\log \gamma/C$  at the two temperatures are 4.4 and 3.7, or about four times the value for hemoglobin in glycine at 25°.

These figures may be compared with the theoretical values for the limiting slope,  $K_R^* - K_S^*$ , calculated for the interaction of lactoglobulin with glycine following the treatment of Fuoss, as described in Chapter 11. This yields, at 5°,  $K_R^* = 2.26$ ,  $K_S^* = 0.48$ ,  $K_R^* - K_S^* = 1.78$ ; and, at 25°,  $K_R^* = 2.06$ ,  $K_S^* = 0.44$ ,  $K_R^* - K_S^* = 1.62$ . As in the case of hemoglobin, the calculated values are too small, although of the correct order of magnitude.

The similarity in the action of sodium chloride and of glycine is evident, both in the interactions of dipolar ions and of ions and dipolar ions; and  $-\log \gamma$  would appear to increase by something less than the first power of the dipole moment. From measurements in Fig. 13 it might be deduced that the order in which ions and other dipolar ions increase the solubility of a protein is the same, but this cannot yet be considered a general rule, for certain proteins, readily dissolved by neutral salts, are not appreciably dissolved by amino acids such as glycine.

### X. INTERACTIONS IN PROTEIN SOLUTIONS CALCULATED FROM ELECTROMOTIVE FORCE AND OSMOTIC MEASUREMENTS\*

The interaction of a soluble protein with an electrolyte may be determined from the electromotive force of a cell in which the electrolyte is transferred from a solution in which its concentration is  $m_3$  and the protein concentration is zero, to a solution in which its concentration is still  $m_3$  and the protein concentration is  $m_2$ . The electromotive force of such a cell is proportional to  $\log \gamma_3/\gamma_3^0$ , the logarithm of the ratio of the mean activity coefficient of the ions in the solution containing protein to that in the protein free solution. Joseph<sup>80</sup> has studied such a cell for zinc chloride in isoionic gelatin and expresses his results as

$$-\log \gamma_3/\gamma_3^0 = (am_3^{-1/2} - b)m_2 - cm^{-1/2}m_2^2 \quad (15)$$

which, transformed by means of equation 16, Chapter 11 (p. 260), yields:

$$-\log \gamma_2/\gamma_2^0 = 2\nu(a - 2cm_2)m_3^{1/2} - \mu bm_3 \quad (16)$$

with  $a = 10.5$ ,  $b = 5.3$ , and  $c = 880$ .

\* We are indebted to George Scatchard for writing the major part of section X.  
<sup>80</sup> Joseph N. R., *J. Biol. Chem.*, 116, 353 (1936)

most seed proteins by acids depends largely on the presence of mineral salts in their solutions." (p. 58) Similarly Hardy<sup>13</sup> showed clearly that the addition of salt to serum globulin shifts the point of minimum solubility to a more acid range, the magnitude of the shift depending on the salt employed, as well as on its concentration. In summarizing the combined effects of salts, acids and alkalis on serum globulin, he concluded, "One feature of fundamental importance, which is never obscured . . . is the antagonism between the solvent actions of salts and acids, and the additive nature of the combined solvent action of salts and alkali."<sup>13</sup>, p. 323 This statement exactly characterized the behavior of many other proteins in acids and alkaline salt solutions. Thus myosin<sup>14</sup>, in solutions alkaline to pH 6.0, is dissolved by both salts and alkalis, the solvent actions of the two tending to reinforce each other; from about pH 4.7 to 6.0, myosin is virtually insoluble at any salt concentration; at reactions acid to pH 4.6 it forms a clear solution, which, however, is precipitated by even small additions of neutral salts.

Thus the "antagonism between the solvent action of salts and acids" on proteins, long ago pointed out by Osborne and by Hardy, appears to be a wide-spread phenomenon, found in proteins in very various types. It may be due to the formation of insoluble acid salts of these proteins, as Osborne suggested for the vegetable globulins; but this interpretation cannot be regarded as established until the chemical composition of the precipitated protein phase in such systems has been accurately determined. When a new saturating body of protein combined with ions is present, the number of components is greater and equations more complex than those thus far developed must be employed to describe the phenomena.

#### VI. SOLUBILITY IN SYSTEMS CONTAINING INSOLUBLE PROTEIN SALTS

Proteins do in fact form insoluble salts both with anions and cations. The vegetable globulin, edestin, as Osborne demonstrated in 1902<sup>14a</sup>, is not only relatively insoluble itself, but forms a relatively insoluble hydrochloride. This is unquestionably one of the causes of the wide precipitation zone of this protein. Rona and Michaelis<sup>15</sup> estimated from the maximum precipitation of edestin in a series of phosphate buffers, that its isoelectric point corresponded to a hydrogen ion concentration of  $1.3 \times 10^{-7}$ . Later Michaelis and Mendelssohn<sup>16</sup> using acetate buffers in which to study the optimum flocculation of edestin believed that the isoelectric point coincided with a hydrogen ion concentration of  $2.5 \times 10^{-6}$ . Hitchcock<sup>17</sup> noted that the point of maximum precipitation of the preparation of edestin

<sup>13</sup> Hardy, W. B., *J. Physiol.*, 33, 251 (1905).  
<sup>14</sup> Osborne, T. B., *Biol. Chem.*, 80, 980 (1930). Also unpublished observations.

He has also made measurements with other isoionic proteins and other electrolytes<sup>81</sup>, which he expresses with a simplified equation in which  $b$  and  $c$  are taken as zero. The values of  $a$  thus obtained are given in Table 8. For the interaction with calcium chloride—the only salt studied with all the proteins investigated— $a$  increases in the order: serum albumin, carboxyhemoglobin, gelatin, pseudoglobulin. The molal dielectric increments of these proteins increase in the same order.

The only EMF data directly comparable with the available solubility measurements are those on carboxyhemoglobin in sodium chloride. Here the two methods give results of the same order of magnitude. Those obtained by the electromotive force method "are, however, based on EMF of about one millivolt, and are not of great significance. . . . The results for the other proteins are more significant, since for all of them potentials of at least 3 or 4 millivolts could be obtained"<sup>81, p. 400</sup>.

TABLE 8. INTERACTIONS OF PROTEINS AND SALTS FROM ELECTROMOTIVE FORCE MEASUREMENTS

	Gelatin	Protein Carboxy- hemoglobin	Serum Albumin	Pseudo- globulin
Molecular weight.....	61,500	67,000	70,000	159,000
Electrolyte.....		$a$ [Equation 15]		
ZnCl <sub>2</sub> .....	8.5			
CaCl <sub>2</sub> .....	5.3	2.5	2.0	9.8
MgCl <sub>2</sub> .....	5.6			
NaCl.....		1.1		6.6

The interaction may also be determined by the effect of the protein on the distribution of the electrolyte across a semipermeable membrane, as discussed in Chapter 17. This method could also be extended to non-electrolytes but no measurements are now available. From Northrop and Kunitz's measurements<sup>82</sup> of the effect of isoionic gelatin on the distribution of certain electrolytes, Joseph<sup>80</sup> has calculated  $a$  and  $b$  with equation 15 simplified by taking  $c$  equal to zero. His results are given in Table 9. The results of Scatchard, Batchelder and Brown<sup>83</sup> for the membrane equilibrium of sodium chloride with serum albumin over a wide range of pH are discussed in Chapter 17. By the simplified equation 16 with  $b$  and  $c$  zero, they yield about two for  $a$  for albumin with valence from +10 to -25, but this value is tripled for valence +27.

Protein-electrolyte interaction may also be calculated from the effect of electrolytes on the osmotic pressure of the proteins. Northrop and Kunitz<sup>84</sup> have measured the effects of a number of salts on the osmotic pressure of 10% isoionic gelatin solutions. AlCl<sub>3</sub>, NaCNS, and perhaps

<sup>81</sup> Joseph, N. R., *J. Biol. Chem.*, **126**, 389 (1938).

<sup>82</sup> Northrop, J. H., and Kunitz, M., *J. Gen. Physiol.*, **11**, 481 (1927).

<sup>83</sup> Scatchard, G., Batchelder, A. C., and Brown, A., Buffalo Meeting of the American Chemical Society, Sept. 10, 1942.

<sup>84</sup> Northrop, J. H., and Kunitz, M., *J. Gen. Physiol.*, **8**, 317 (1926).

$\text{Li}_2\text{SO}_4$  show specific effects in dilute solutions. The pressures with other salts show an initial increase proportional to the square root of the ionic strength, pass through a maximum and then decrease almost linearly in the ionic strength. The initial increase has the same proportionality factor for all the salts, but the decrease shows a typical Hofmeister series:  $\text{MgCl}_2$ ,  $\text{CaCl}_2$ ,  $\text{LiCl}$ ,  $\text{NaI}$ ,  $\text{NaBr}$ ,  $\text{NaNO}_3$ ,  $\text{NaClO}_3$ ,  $\text{NaCl}$ ,  $\text{KCl}$ ,  $\text{Na acetate}$ .

The interpretation of these results cannot be simple. Even at the isoionic point the proteins exist partly as ions, so there should be a term in the square root of the ionic strength which explains about half the effect of the alkali and alkaline earth salts on osmotic pressure of gelatin, and about half their interactions with carboxyhemoglobin and serum albumin. The theoretical value for the logarithm of the activity of a dipolar ion, calculated from equation 10, plotted against the square root of the ionic strength has a zero initial slope, but has an inflection so that it is nearly linear over most of the experimental range.\* Perhaps that part of the effect which depends upon the ionic strength may be attributed to the

TABLE 9. INTERACTION OF ISOIONIC GELATIN AND SALTS FROM MEMBRANE EQUILIBRIA

Electrolyte	a	b
$\text{ZnCl}_2$ .....	10.7	5.3
$\text{MgCl}_2$ .....	6.9	1.5
$\text{CuCl}_2$ .....	37.5	50
$\text{LaCl}_3$ .....	24.2	6.2
$\text{AlCl}_3$ .....	46.8	75

electrostatic effects considered in equation 10. The remainder should probably be attributed to the specific short range effects which we call chemical. The chemical effects are very important for the heavy metals listed in Table 9. Little more can be said from the limited results now available.

The measurements of osmotic pressure can also give the protein-protein interaction, as indicated in Chapter 17. The fact that, for the proteins thus far investigated, the osmotic pressure-concentration ratio always increases, shows that these isoionic proteins increase their own activity coefficients, though the increase is sometimes very small. It may be affected greatly by the addition of a salt or a non-electrolyte. The osmotic pressure-concentration ratio for 10% gelatin is about twice the value at zero concentration. The maximum values with uni-univalent salts are about four times the value at zero concentration, so the effect of gelatin on its own  $\log \gamma$  is about tripled. On the other hand the effect for isoionic serum albumin is little influenced by the addition of salt. The derivative

\* That the effect of chemical combination behaves similarly is well known from the fact that adsorption measurements may be fitted over a considerable range by either the Freundlich equation with exponent one-half or by the Langmuir equation.

$d \ln \gamma_2/dm_2$  is about  $-2000$  when the valence is  $27$ , has a maximum of  $900$  when the valence is about  $-5$  and falls to  $-250$  for a valence of  $-25$ . There are not yet sufficient results to warrant an attempt at explanation.

Hitherto most of the data concerning interactions between proteins, salts and non-electrolytes have been derived from solubility measurements. The electromotive force and osmotic methods discussed in this section have been employed by very few investigators; largely perhaps because the measured effects are small, and great precision in the measurements is required if significant results are to be obtained. In future these methods will probably be far more widely used. This is true partly because they furnish a valuable independent estimate of the activity coefficients deduced from solubility measurements. They have, moreover, the great advantage of permitting independent variation of the salt and the protein concentration in the system. Measurements may be carried out, in a given solvent, at any desired protein concentration up to saturation. It is thus possible to study certain most important interactions, namely, those between proteins, which are far more difficult to study by solubility measurements. A few examples of protein-protein interaction have been revealed by the experiments discussed above; further studies in this field, utilizing all the experimental techniques available, should lead to definite and comprehensive knowledge of the interaction of protein molecules with other ions and dipoles, and with each other. Such knowledge forms an essential foundation for a deeper understanding of physiological systems.



## Chapter 25

# The Theory of Electrophoretic Migration

BY HANS MUELLER

The behavior of the electrically charged micelles of colloids and proteins in an electric field is similar to that of the ions in an electrolyte. Depending on the sign of their charge, they migrate toward either the cathode or the anode. Since in solutions which contain a number of different colloidal components the various types of particles usually travel with different speeds, the electrophoretic migration offers a method for the separation of these components. The quantitative analysis of proteins by means of this method, as developed by Tiselius<sup>1</sup>, Theorell<sup>2</sup>, Longworth<sup>3</sup> has become of great practical significance, but the theory of the phenomenon has not yet found a completely satisfactory formulation. Measurements of the migration speed have shown that it varies with the pH and the activity of the electrolyte, and in some cases also with the size and shape of the particles<sup>4-13</sup>. The attempts to account for these variations are carried out in two successive steps. The first step consists in establishing a correlation between the migration speed and the electrical properties of the micelle, *i.e.*, its charge or potential. The second step deals with the dependence of these quantities on the pH and the activity of the solution and the size of the particle. The latter problem having been discussed in Chapter 20 we can here limit our consideration to the theory of the migration speed.

A uniform electric field of intensity,  $E$ , exerts on an electrically charged particle a force  $F = QE$ . On the basis of this elementary law it should, therefore, be possible to determine the charge  $Q$  of a colloidal particle from measurements of the force  $F$  in an electric field of given intensity. This indeed is the method used by Millikan in his famous oil-drop experiment

<sup>1</sup> Tiselius, A., Dissertation, Upsala (1930); *Nova Acta Reg. Soc. Sci., Upsala*, (IV) 7, 4 (1930); *Trans. Far. Soc.*, 33, 524 (1937).

<sup>2</sup> Theorell, H., *Biochem. Z.*, 275, 1 (1934).

<sup>3</sup> Longworth, L. G., *J. Am. Chem. Soc.*, 61, 520 (1939); Longworth, L. G., and MacInnes, D. A., *Chem. Rev.* 24, 271 (1939); *J. Am. Chem. Soc.*, 62, 705 (1940).

<sup>4</sup> Picton, H., and Linder, S. E., *J. Chem. Soc.*, 71, 588 (1957).

<sup>5</sup> Hardy, W. B., *J. Physiol.*, 24, 153, 295, (1900); 33, 25, (1905).

<sup>6</sup> Loeb, J., *J. Gen. Physiol.*, 5, 385 (1921).

<sup>7</sup> Michaelis, L., *Die Wasserstoffionenkonzentration*. Berlin (1914) (1922).

<sup>8</sup> Kemp, I., and Rideal, E. K., *Proc. Roy. Soc. (London)* A147, 11 (1934).

<sup>9</sup> Smith, E. R. B., *J. Biol. Chem.*, 108, 187, 135 (1935); 113, 473 (1936).

<sup>10</sup> Abramson, H. A., "Electrokinetic Phenomena and Their Applications to Biology and Medicine," New York, Chemical Catalog Co., Inc., Reinhold Publishing Corp., 1934; Abramson, H. A., Gorin, M. H. and Moyer, L. S., *Chem. Rev.*, 24, 345 (1939). Abramson, H. A., "Electrophoresis of Proteins," New York, Reinhold Publishing Corporation, 1942.

<sup>11</sup> Tiselius, A., *Biochem. J.*, 31, 313 (1937); Tiselius, A., and Svensson, H., *Trans. Far. Soc.*, 36, 16 (1940).

<sup>12</sup> Davis, B. D., and Cohn, E. J., *Ann. N. Y. Acad. Sci.*, 39, 27, (1939); *J. Am. Chem. Soc.*, 61, 2092 (1939).

<sup>13</sup> Adair, G. S., and Adair, M. E., *Trans. Far. Soc.*, 36, 23, (1940).

for the determination of the electronic charge. The force  $F$  is measured by balancing it with a second force,  $F'$ , so that the resulting force,  $F - F'$ , vanishes. According to Newton's first law the particle will then either stay at rest or move with a constant velocity.

The electrokinetic investigations, in common with Millikan's experiment, employ as balancing force the friction which is exerted on a moving particle by the surrounding viscous medium. If this medium is a perfect insulator, containing a negligible amount of free ions, and if the particle is spherical of radius  $a$ , the friction force is given by Stokes' law as

$$F' = 6\pi\eta Va \quad (1)$$

where  $V$  is the migration velocity and  $\eta$  is the coefficient of viscosity of the medium. If the above conditions are met, as they are in the oil-drop experiment, the balancing condition  $F = F'$  furnishes for the mobility  $u$

$$u = V/E = Q/6\pi\eta a \quad (2)$$

Inasmuch as the radius,  $a$ , of small particles is seldom accurately known, it is convenient to eliminate it from the equation by introducing the electric potential. The potential of a spherical particle in an insulating medium with the dielectric constant  $D$  is

$$\psi_0 = Q/Da \quad (3)$$

Hence equation 2 may be written

$$u = \psi_0 D / 6\pi\eta \quad (4)$$

This simple relation between the mobility and the electric potential of spherical particles in an insulating medium involves only the dielectric constant and the viscosity of the medium.

Relations analogous to equation 4 can readily be given for particles of other shapes. For instance, for rod-shaped micelles of length  $2a$  with circular cross-section of radius  $b \ll a$  the viscous resistance for migration in direction along the rod is

$$F' = 4\pi\eta Va / \ln (2a/b)$$

while for crosswise migration

$$F' = 8\pi\eta Va / \ln (2a/b)$$

For these particles the relation between charge and potential is

$$\psi_0 = \frac{Q}{aD} \ln (2a/b)$$

whence  $u = \psi_0 D / 4\pi\eta$  for lengthwise, and  $u = \psi_0 D / 8\pi\eta$  for crosswise migration of rod-shaped particles. For particles of arbitrary shape the migra-

tion speed in an insulating medium is always given by an equation of the form

$$u = \psi_0 D / C \eta$$

The value of the numerical factor  $C$  varies with the shape of the micelle and also with the orientation of the particle relative to the direction of migration.

Inasmuch as the micelles of colloids and proteins are surrounded by an electrolytically conductive medium, the essential requirements for the applicability of Stokes' law and hence also of equation 4 are not satisfied for the migration of these particles. As was pointed out in Chapter 3, the electrostatic forces between the charge of the particle and the ions of the electrolyte give rise to the formation of a diffuse electric double layer around the micelle. The presence of this ionic atmosphere alters the above considerations in two respects. First, it lowers the value of the potential from  $Q/Da$  to

$$\psi_0 = \frac{Q}{Da} - \psi' \quad (3')$$

where the reduction  $\psi'$  depends on the constitution of the double layer. If  $\psi_0$  is sufficiently small,  $\psi'$  can be calculated from Debye-Hückel's theory. This theory gives  $\psi_0 = Q/Da(1 + \kappa a)$ , whence  $\psi' = Q\kappa/D(1 + \kappa a)$ .

The second change arises from the fact that the applied field,  $E$ , acts also upon the ions of the double layer. The forces which the field exerts on the ions are transmitted to the liquid and alter its flow. Since the charge of the ion cloud has the opposite sign to the charge of the particle, the electrically induced additional flow is in the direction opposite to the motion of the particle. Consequently the friction force is increased and Stokes' law for spherical particles must be replaced by an equation of the form

$$F' = 6\pi\eta Va + qE \quad (1')$$

The additional "electrophoretic" force  $F_e = qE$  will be proportional to the field intensity  $E$ , and the factor  $q$  depends on the constitution of the double layer. By introducing 1' and 3' into the balancing condition  $QE = F'$ , one finds now

$$u = \frac{\psi_0 D}{6\pi\eta} + u' \quad (4')$$

where

$$u' = (\psi' D - q/a) / 6\pi\eta \quad (5)$$

Due to the increased friction, spherical particles of equal charge and radius will migrate more slowly in a conductive solution than in an insulating

medium of the same viscosity. However, the above equations are not sufficient to predict how the mobility of particles with the same potential,  $\psi_0$ , will differ in the two kinds of media. Since  $\psi'$  and  $q$  have the same sign, both being positive for positively charged particles, it is impossible to decide, without further calculations, whether the mobility  $u'$  is positive, negative or zero. Inasmuch as the migration in an insulating medium represents the limiting case of electrophoretic migration in an electrolyte of infinitely small concentration, it is obvious that  $u'$  tends asymptotically toward zero with diminishing electrolyte concentration. Equation 4 therefore is valid in the limiting case of particles with a double layer whose thickness is much larger than the radius  $a$  of the particle. Hence it is applicable to the electrophoretic migration of ions and very small particles, provided the so-called relaxation effect of the ionic atmosphere is negligible. This conclusion is in accordance with Debye-Hückel's theory for the mobility of ions.

In most colloids and proteins the thickness of the double layer is smaller than, or of the same order of magnitude as, the radius of the particles. In these cases the correction terms  $\psi'$  and  $F_e$  are of the same order of magnitude as  $Q/Da$  and  $6\pi\eta Va$ , respectively, and there is therefore no reason to expect that  $u'$  vanishes. To find the mobility equation for larger particles and to decide the range of applicability of equation 4 it is necessary to study the hydrodynamics of the electrophoretic flow.

### The Electrophoretic Flow Parallel to Plane Surfaces

Since electrophoretic migration proceeds with small velocities, it is governed by the laws of the hydrodynamics of laminar flow in viscous liquids. The simplest example of this kind of flow is that of a liquid between two large parallel plates, one being at rest, while the other is translated with a constant velocity  $V$ . The theory assumes that at the surfaces of the plates the liquid adheres to the surface and moves with the same velocity as the plates. If  $z$  designates the direction normal to the plates and  $x$  points in the direction of  $V$ , the velocity  $v_x$  of the liquid at the distance  $z$  from the stationary plate is

$$v_x(z) = Vz/L \quad (6)$$

where  $L$  is the distance between the two plates. The fundamental law of viscosity states that this flow exerts on a unit area of the plates a friction force

$$dF/dA = -\eta V/L = -\eta dv_x/dz \quad (7)$$

where  $\eta$  is the coefficient of viscosity of the liquid.

Let us now consider the flow created by the motion of a large thin plane flake moving midway between two parallel stationary plates with a separation  $2L$ . The flake is assumed to be of infinite area,  $A$ , and infinitely thin so that edge effects are negligible. The flake carries a charge,  $Q$ , which is uniformly distributed over its surface, giving rise to a charge density on the surface of  $\sigma = Q/A$ . The motion is produced by a uniform field,  $X$ , parallel to the surface. The fixed plates shall be at potential zero.

If the liquid is an insulator with a dielectric constant  $D$ , the flake and the plates form a parallel plate condenser and the flake is at a potential  $\psi_0 = 4\pi\sigma L/D$ . The flow on both sides of the flake is given by equation 6 and the friction force is, according to 7,  $F' = \eta AV/L$ . This force balances the electric force  $QE' = A\sigma X$ , whence  $V = \frac{\sigma XL}{\eta} = \psi_0 DX/4\pi\eta$ .

If the liquid is an electrolyte the potential gradient between flake and plates is not uniform, as it is in the insulating liquid. Whatever the constitution of the double layer may be, provided only that its thickness be small in comparison to the plate separation  $L$ , the potential  $\psi$  will decrease very rapidly from its value  $\psi_0$  at the surface of the particle to zero in the solution outside the double layer and at the plates. The unequal distribution of the positive and negative ions within the double layer gives rise to a space charge of density  $\rho$ . This charge density is related to the potential variation according to Poisson's equation of electrostatics

$$\frac{d^2\psi}{dz^2} = -4\pi\rho/D \quad (8)$$

At the surface of the flake it follows from Gauss' theorem

$$\sigma = -\frac{D}{4\pi} (d\psi/dz)_0 \quad (9)$$

where  $(d\psi/dz)_0$  is the potential gradient at the surface.

To arrive at the equation for the electrophoretic flow around the flake, we consider a small volume element  $d\tau = dx \cdot dy \cdot dz$  of the liquid at an arbitrary distance,  $z$ , from the flake. If  $v_x(z)$  denotes the unknown velocity of the liquid, we find from equation 7 that the flow exerts on this element a friction force

$$dF_x^v = [-\eta(dv_x/dz)_z + \eta(dv_x/dz)_{z+dz}]dx \cdot dy = \eta d^2v_x/dz^2 d\tau \quad (10)$$

because  $-\eta(dv_x/dz)_z dx \cdot dy$  is the friction force on the lower surface at  $z$  and  $\eta(dv_x/dz)_{z+dz} dx \cdot dy = \eta[(dv_x/dz)_z + d^2v_x/dz^2 \cdot dz]dx \cdot dy$  is the force on the upper surface of area  $dx \cdot dy$ .

In addition to the friction force  $dF_x^v$ , an electric force  $dF_x^e$  is acting on the liquid in the element  $d\tau$ , because it contains a small amount of charge

$dQ = \rho d\tau$  which is under the influence of the applied field  $X$ . The force  $dQ \cdot X$  on the ions is transmitted to the liquid, whence, according to 8,

$$dF_z^{e1} = -\frac{DX}{4\pi} d^2\psi/dz^2 \cdot d\tau \quad (11)$$

Since the flow is stationary, the element  $d\tau$  moves with constant velocity. Hence the total force  $dF_z^v + dF_z^{e1}$  must be zero, and this condition leads, according to 10 and 11, to the differential equation for  $v_z$

$$d^2 v_z/dz^2 = \frac{DX}{4\pi\eta} d^2\psi/dz^2 \quad (12)$$

which has the solution

$$v_z = \frac{DX}{4\pi\eta} \psi + az + b$$

The values of the integration constants are  $a = b = 0$  because the boundary conditions at the plates, where  $\psi = 0$ , require  $v_z(\pm L) = 0$ . From the boundary condition at the surface of the flake, where  $\psi = \psi_0$ ,  $v_z = V$ , it follows that

$$V = \frac{DX}{4\pi\eta} \psi_0 \quad (13)$$

The friction force exerted on the flake is, according to 7 and 9

$$F' = -A\eta(dv_z/dz)_0 = -\frac{ADX}{4\pi\eta} (d\psi/dz)_0 = A\sigma X$$

*i.e.*, it is of the same magnitude as the electric force  $QX$  on the charge of the flake. Hence the flake moves with the constant velocity  $V$  given by equation 13.

Although the electrophoretic mobility equation 13 for the flake is identical with that for its mobility in an insulating liquid, it should be noted that the flow of the liquid is entirely different in the two cases. In the insulating liquid the flow extends over the whole volume and depends on the plate distance  $L$ . In the electrophoresis the flow is confined to the narrow range of the double layer; outside this range the liquid stays at rest and the plate distance does not affect the flow if it is much larger than the thickness of the double layer. It is therefore natural to expect that in general, *i.e.*, for other forms of the particle, the mobility equations are different for the migration in insulating liquids and for electrophoretic migration.

Equation 13 was first given by Smoluchowski and is frequently referred to as Smoluchowski's mobility equation. A generalization of the above derivation shows that it is applicable to all cases where the flow of the liquid proceeds along straight lines. It holds therefore also for the migra-

tion of long rod-shaped particles if they are moving lengthwise and for the electroösmosis in straight capillaries. The equation is independent of the structure of the double layer. While Smoluchowski assumes that his equation was applicable to particles of all shapes and was not affected by the orientation and size of the micelles, the more recent developments of the theory indicate that this is true only with the reservation that the thickness of the double layer is small as compared to the linear dimensions of the particle.

The differential equation 12 for parallel flow must be supplemented by an additional term if the flow creates pressure differences in the liquid. The influence of pressure may be discussed for the electroösmosis in a flat, closed cell of the type commonly used in the microscopic method of electrophoretic investigation. If the distance between the windows of such a cell is small in comparison to its length and width, the electroösmotic flow is practically identical with that between two parallel plates. If the cell were open on both ends the liquid would flow everywhere in the same direction. In a closed cell, however, the unidirectional flow creates a pressure difference between the ends of the cell, and, since the liquid is incompressible, a counterflow through the middle of the cell is set up.

If the plate distance between the windows is again  $2L$  in the  $z$  direction, the origin chosen in the midpoint, and if the field  $X$  and the flow velocity are oriented along the  $x$  axis, the pressure  $p$  will be a function of  $x$  only. The pressure gradient gives rise to an additional force  $dF_x^p$  acting on a volume element,  $d\tau$ , of the liquid, namely

$$dF_x^p = [p(x) - p(x + dx)] dy \cdot dz = -\frac{dp}{dx} d\tau \quad (14)$$

The condition for stationary flow is now  $dF_x^v + dF_x^{e1} + dF_x^p = 0$  and leads, according to 10, 11 and 14, to the differential equation

$$\eta \frac{d^2 v_x}{dz^2} - \frac{dp}{dx} = \frac{DX}{4\pi} \frac{d^2 \psi}{dz^2} \quad (15)$$

Since  $v_x$  and  $\psi$  are independent of  $x$ , partial differentiation with respect to  $x$  gives  $\partial^2 p / \partial x^2 = 0$ , whence  $p = p_0 + cx$ ,  $dp/dx = c$ . The solution of 15 is, therefore,

$$v_x = \frac{DX\psi}{4\pi\eta} + \frac{cz^2}{2\eta} + bz + V_0$$

To evaluate the integration constants we note that the electroösmotic flow is the result of a potential difference  $\psi_0$  between the cell walls and the liquid. Hence for  $z = \pm L$ ,  $v_x = 0$ ,  $\psi = \psi_0$  and, therefore,  $b = 0$  and

$$-\frac{DX\psi_0}{4\pi\eta} = \frac{cL^2}{2\eta} + V_0 \quad (16)$$

A second relation between  $c$  and  $V_0$  follows from the requirement that the total amount of liquid flowing through a cross-section of the cell must be zero. This condition yields

$$\int_{-L}^L v_x dz = \frac{DX}{4\pi\eta} \int_{-L}^L \psi dz + \frac{cL^3}{3\eta} + 2V_0L = 0$$

In all practical cases the thickness of the double layers along the upper and lower window is so small that the transport of liquid in the double layer, represented by the first term, is negligible. The above equation gives therefore

$$V_0 = -cL^2/6\eta$$

where  $V_0$  is the flow velocity in the middle of the cell. From 16 one finds now for the pressure gradient

$$c = dp/dx = -3DX\psi_0/4\pi L^2$$

and the velocity distribution is

$$v_x = \frac{DX}{4\pi\eta} \left[ \psi - \frac{1}{2}\psi_0 \left( 3 \frac{z^2}{L^2} - 1 \right) \right] \quad (17)$$

Throughout the cell, except near the walls,  $\psi = 0$ . Hence the velocity of the bulk of the liquid is given by

$$v_x = -\frac{DX\psi_0}{8\pi\eta} \left( 3 \frac{z^2}{L^2} - 1 \right) \quad (17')$$

We note in particular that  $v_x = 0$  when  $z = L/\sqrt{3}$ . This fact plays an important role in the evaluation of the electrophoretic measurements carried out by the microscopic method in closed, flat cells. The particles which are suspended in the moving liquid of these cells possess a velocity which is the algebraic sum of the velocity  $v_x$  of the liquid and the migration velocity of the particle relative to the liquid. The latter, *i.e.*, the true electrophoretic speed of the particle, is obtained directly when the observations are made at a depth where the liquid is at rest. In accordance with the above consideration the measurements<sup>14</sup> confirm that this occurs

at the distance  $0.211d = \frac{d}{2} \left[ 1 - \frac{1}{\sqrt{3}} \right]$  from either wall, when  $d = 2L$  is the thickness of the cell.

In evaluating the electric force  $dF^{e1}$  on the volume element  $d\tau$  only the action of the applied field  $X$  was considered. Actually, however, a second electric intensity  $Z = -d\psi/dz$ , due to the field of the double layer, also is acting upon the charges  $dQ = \rho d\tau$ . This second field creates a force

$$F_z^{e1} = \frac{D}{4\pi} \frac{d^2\psi}{dz^2} \cdot \frac{d\psi}{dz} d\tau = \frac{D}{8\pi} \frac{d}{dz} \left( \frac{d\psi}{dz} \right)^2 d\tau$$

<sup>14</sup> Abramson, H. A., "Electrokinetic Phenomena and Their Application to Biology and Medicine," p. 71. New York, Chemical Catalog Co., Reinhold Publishing Corp., 1934



which is directed normal to the charged plane surface of either the flake or the walls of the flat cell. This force cannot influence the tangential flow of the liquid and can therefore be neglected in the theory of electrophoretic migration. It gives rise to a pressure-gradient normal to the plane

$$dp/dz = \frac{D}{8\pi} \frac{d}{dz} \left( \frac{d\psi}{dz} \right)^2 \quad (18)$$

This gradient exists also when the particle is at rest and it is not changed by the movement of the liquid. It gives rise to a pressure difference so that the pressure exerted on a charged colloidal particle is larger than the ordinary hydrostatic pressure in the liquid. From 18 it follows that the additional pressure,  $p_0$ , on a plane surface is

$$p_0 = \frac{D}{8\pi} \left( \frac{d\psi}{dz} \right)_0^2 = \frac{2\pi\sigma^2}{D}$$

On the basis of a more complete calculation by Dällenbach<sup>15</sup> and Zwicky<sup>16</sup>, which takes into account the electric forces on the ions and on the induced polarization charges, Mueller<sup>17</sup> finds for the additional pressure

$$p_0 = 2\pi\sigma^2 \left[ \frac{(D-1)(3D+7)}{5D^2} + 1/D^2 \right] \quad (19)$$

In aqueous solutions ( $D = 80$ ) this pressure is, approximately,  $p_0 = 6\pi\sigma^2/5$ , and, since the surface charge  $\sigma$  of many proteins and colloids is between  $10^3$  and  $10^4$  esu, the pressure  $p_0$  may reach values between 10 to 100 atmospheres. In the double layer around ions this pressure amounts to several thousands of atmospheres and, as Zwicky<sup>16</sup> has shown, plays an important role in the theory of the specific heat of electrolytes. Very little is known whether or not this pressure effect is of importance in the physics of colloids and proteins. Mueller<sup>17</sup> has pointed out that it may contribute to the understanding of the problem of the stability of colloidal solutions. It seems unlikely that the pressure influences the migration of particles, provided that it does not alter the viscosity and dielectric constant of the liquid.

### The Electrophoretic Migration of Spherical Particles

In contrast to the electrophoretic flow along plane surfaces the problem of electrophoretic migration of colloidal particles offers great mathematical difficulties. To be sure, the differential equation for the flow about micelles of arbitrary shapes can readily be obtained by a generalization of equation 15, as follows. The velocity,  $v$ , of an arbitrary volume element,  $d\tau$ ,

<sup>15</sup> Dällenbach, W., *Physik. Z.*, **27**, 632 (1926).

<sup>16</sup> Zwicky, F., *Physik. Z.*, **27**, 271 (1926).

<sup>17</sup> Mueller, H., *J. Phys. Chem.*, **39**, 743 (1935); *Cold Spring Harbor Symp. Quant. Biol.*, **1**, 60 (1933).

of the liquid, located at a position given by the coördinates  $x, y, z$ , has three components  $v_x, v_y, v_z$ . The flow creates a pressure distribution  $p(x, y, z)$  and the electric field, arising from the externally applied voltage and from the charges of the particle and the ions in the double layer, is given by a potential  $\Phi(x, y, z)$ . Each one of the three forces  $dF^v, dF^{e1}$  and  $dF^p$  on a volume element has three components  $dF_x, dF_y$  and  $dF_z$ . In the following we shall give only the  $x$  components of these forces, because the analogous equations for the  $y$  and  $z$  components are obtained simply by a permutation of the indices  $x, y, z$ .

Application of equation 7 to all six sides of the rectangular volume element yields

$$dF_x^v = \eta \left[ \nabla^2 v_x - \frac{\partial}{\partial x} \left( \frac{\partial v_x}{\partial x} + \frac{\partial v_y}{\partial y} + \frac{\partial v_z}{\partial z} \right) \right] d\tau \quad (10')$$

The pressure gradient creates a force

$$dF_x^p = - \frac{\partial p}{\partial x} d\tau \quad (14')$$

and the electric force is  $dF_x^{e1} = \rho X d\tau$ . Since the components of the field intensity are  $X = - \frac{\partial \Phi}{\partial x}$ , and the charge density  $\rho$  is given by Poisson's law to  $\rho = - \frac{D}{4\pi} \nabla^2 \Phi$ , the electric force is

$$dF_x^{e1} = \frac{D}{4\pi} \nabla^2 \Phi \frac{\partial \Phi}{\partial x} d\tau \quad (11')$$

Finally the velocity  $v$  must satisfy the continuity equation

$$\frac{\partial v_x}{\partial x} + \frac{\partial v_y}{\partial y} + \frac{\partial v_z}{\partial z} = 0 \quad (20)$$

which states that the amount of liquid in the element  $d\tau$  remains constant and which is a consequence of the negligible compressibility of the liquid.

According to 10', 14', 11' and 20, the total force on a volume element has the components

$$dF_x = \left[ \eta \nabla^2 v_x - \frac{\partial p}{\partial x} + \frac{D}{4\pi} \nabla^2 \Phi \frac{\partial \Phi}{\partial x} \right] d\tau$$

This force is not zero, except in the case of parallel flow, where the velocity of any volume element remains constant. In the flow around a particle the velocity of the liquid changes its magnitude and direction; hence the mass  $dm$  of the liquid in  $d\tau$  has an acceleration, and the equation of motion is therefore  $dF_x = a_x dm$ . For the slow motion of small particles in a steady field Lamb and Oseen<sup>18</sup> have shown that the inertia term  $a \cdot dm$  is

<sup>18</sup> Lamb, H., *Hydrodynamics*, p. 608, 8th edition, Cambridge University Press, 1932.

small in comparison to the other forces and can in most cases be neglected. It is therefore permissible to assume  $dF = 0$ , but it should be noted that we thereby limit our consideration to the electrophoresis in a steady electric field. In alternating electric fields the inertia term becomes increasingly more important, the higher the frequency. The condition  $dF = 0$  yields

$$\begin{aligned} -\eta \nabla^2 v_x + \frac{\partial p}{\partial x} &= \frac{D}{4\pi} \nabla^2 \Phi \frac{\partial \Phi}{\partial x} \\ -\eta \nabla^2 v_y + \frac{\partial p}{\partial y} &= \frac{D}{4\pi} \nabla^2 \Phi \frac{\partial \Phi}{\partial y} \\ -\eta \nabla^2 v_z + \frac{\partial p}{\partial z} &= \frac{D}{4\pi} \nabla^2 \Phi \frac{\partial \Phi}{\partial z} \end{aligned} \quad (21)$$

which, in conjunction with equation 20, furnishes four differential equations for the four unknowns  $v_x$ ,  $v_y$ ,  $v_z$  and  $p$ .

In order to arrive at simpler equations, one differentiates the three equations 21 with respect to  $x$ ,  $y$ ,  $z$ , respectively. On account of 20, addition of the results leads to an equation for the pressure

$$\nabla^2 p = \frac{D}{4\pi} \left[ \frac{\partial \Phi}{\partial x} \cdot \frac{\partial \nabla^2 \Phi}{\partial x} + \frac{\partial \Phi}{\partial y} \cdot \frac{\partial \nabla^2 \Phi}{\partial y} + \frac{\partial \Phi}{\partial z} \cdot \frac{\partial \nabla^2 \Phi}{\partial z} + (\nabla^2 \Phi)^2 \right] \quad (22)$$

In place of the velocity vector  $v$  it is convenient to introduce the vorticity vector  $w$ , the components of which are defined by

$$w_x = \frac{\partial v_z}{\partial y} - \frac{\partial v_y}{\partial z} \quad (23)$$

If now the second equation 21 is differentiated with respect to  $z$ , and the last equation 21 differentiated with respect to  $y$ , subtraction of the results yields

$$\nabla^2 w_x = \frac{D}{4\pi\eta} \left[ \frac{\partial \Phi}{\partial y} \cdot \frac{\partial \nabla^2 \Phi}{\partial z} - \frac{\partial \Phi}{\partial z} \cdot \frac{\partial \nabla^2 \Phi}{\partial y} \right] \quad (24)$$

and permutation of  $x$ ,  $y$ ,  $z$  gives two analogous expressions for  $\nabla^2 w_y$  and  $\nabla^2 w_z$ .

The electrophoretic friction force can be calculated if the potential distribution  $\Phi$  is known, and if one succeeds in finding solutions of the differential equations 22 and 24 which satisfy the boundary conditions. The boundary conditions require that at the surface of the particle the velocity of the liquid be the same as that of the particle and that the liquid is at rest at a large distance from the micelle. The mathematical procedure is simplified by referring all quantities to a system of coordinates which moves with the particle. This change does not affect the differential equations, but the boundary conditions are now:  $v = 0$ , at the surface, and  $v = -V$  at a large distance from the particle, if  $V$  denotes the velocity of the particle.

The use of the first boundary condition implies that the rotational Brownian motion of the particle can be neglected. It is not known what, if any, effect the rotation has on the migration speed. Inasmuch as small particles possess a very large angular velocity of rotational temperature motion, it seems plausible to assume that particles of arbitrary shape behave like spheres, and furthermore, since the speed and the axis of rotation change continually, that the average migration speed is identical with that of a non-rotating sphere. No justification of these assumptions has ever been given.

For the problem of the electrophoretic migration of a sphere, it is convenient to employ polar coordinates  $r, \theta, \varphi$  referred to the center of the sphere with the direction of the migration velocity as polar axis. The symmetry of the problem requires that  $v_r, v_\theta, p$  and  $\Phi$  are independent of  $\varphi$  and that  $v_\varphi = 0$ . This has the consequence that the vorticity vector  $w$  is given by a single component  $w = w_\varphi (w_r = w_\theta = 0)$ . Hence, the number of differential equations is reduced to two. They take the form<sup>19</sup>

$$\begin{aligned} \nabla^2 p &= \frac{1}{r^2} \frac{\partial}{\partial r} \left( r^2 \frac{\partial p}{\partial r} \right) + \frac{1}{r^2 \sin \theta} \frac{\partial}{\partial \theta} \left( \sin \theta \frac{\partial p}{\partial \theta} \right) \\ &= \frac{D}{4\pi} \left[ \frac{\partial \Phi}{\partial r} \frac{\partial \nabla^2 \Phi}{\partial r} + \frac{1}{r^2} \frac{\partial \Phi}{\partial \theta} \frac{\partial \nabla^2 \Phi}{\partial \theta} + (\nabla^2 \Phi)^2 \right] \end{aligned} \quad (22')$$

$$\begin{aligned} \frac{1}{r} \frac{\partial^2 (rw)}{\partial r^2} + \frac{1}{r^2} \frac{\partial}{\partial \theta} \left[ \frac{1}{\sin \theta} \frac{\partial}{\partial \theta} (w \sin \theta) \right] \\ = \frac{D}{4\pi r} \frac{1}{r} \left[ \frac{\partial \Phi}{\partial r} \frac{\partial \nabla^2 \Phi}{\partial \theta} - \frac{\partial \Phi}{\partial \theta} \frac{\partial (\nabla^2 \Phi)}{\partial r} \right]. \end{aligned} \quad (24')$$

The streaming velocity is obtained from the vorticity by the use of equation 23, which takes the form

$$w = \frac{1}{r} \frac{\partial}{\partial r} (rv_\theta) - \frac{1}{r} \frac{\partial v_r}{\partial \theta} \quad (23')$$

and  $v$  has to satisfy the continuity equation

$$\frac{1}{r^2} \frac{\partial}{\partial r} (r^2 v_r) + \frac{1}{r \sin \theta} \frac{\partial}{\partial \theta} (v_\theta \sin \theta) = 0 \quad (20')$$

The boundary conditions are  $v_r = v_\theta = 0$ , for  $r = a$ ; and  $v_r = -V \cos \theta$ ,  $v_\theta = V \sin \theta$ , when  $r$  is very large. The friction force on the particle is obtained from

$$F' = 2\pi a^2 \int_0^\pi (p_{rx})_{r=a} \sin \theta d\theta \quad (25)$$

<sup>19</sup> In carrying out the transformation to polar coordinates it should be noted that  $p$  is a scalar,  $w$  a vector quantity. In the notation of vector analysis we have therefore  $\nabla^2 p = \text{div grad } p$ , but  $\nabla^2 w = \text{curl curl } w$ .

where  $a$  is the radius of the sphere and  $(p_{rx})_{r=a}$  is the component in the direction of motion of the force acting on a surface element of the particle. A lengthy but elementary calculation shows that

$$p_{rx} = \left( -p + 2\eta \frac{\partial v_r}{\partial r} \right) \cos \theta - \eta \left( \frac{\partial v_\theta}{\partial r} + \frac{1}{r} \frac{\partial v_r}{\partial \theta} - \frac{v_\theta}{r} \right) \sin \theta \quad (26)$$

It is convenient<sup>20</sup> to consider the solution of the above boundary value problem as consisting of three parts:

$$\begin{aligned} p &= p^{\text{I}} + p^{\text{II}} + p^{\text{III}} \\ w &= w^{\text{I}} + w^{\text{II}} + w^{\text{III}} \\ v_r &= v_r^{\text{I}} + v_r^{\text{II}} + v_r^{\text{III}} \\ v_\theta &= v_\theta^{\text{I}} + v_\theta^{\text{II}} + v_\theta^{\text{III}} \\ F' &= F^{\text{I}} + F^{\text{II}} + F^{\text{III}} \end{aligned} \quad (27)$$

The index I refers to a particular solution of the inhomogeneous differential equations 22' and 24'. The indices II and III refer to solutions of the homogeneous equations which are obtained from 22' and 24' by setting their right sides equal to zero. Inasmuch as the homogeneous equations relate to the flow about an uncharged particle or about a charged particle in an insulating medium, we are justified in assuming that solution II is given by the relations which lead to Stokes' law, namely

$$\begin{aligned} p^{\text{II}} &= \frac{3V\eta a}{2r^2} \cos \theta \\ v_r^{\text{II}} &= V \cos \theta \left[ \frac{3}{2} \frac{a}{r} - 1 - \frac{a^3}{2r^3} \right] \\ v_\theta^{\text{II}} &= -V \sin \theta \left[ \frac{3}{4} \frac{a}{r} - 1 + \frac{a^3}{4r^3} \right] \end{aligned} \quad (28)$$

and

$$F^{\text{II}} = 6\pi\eta aV.$$

Since this solution satisfies the boundary conditions, the two remaining parts, I and III, have to comply with the much simpler conditions that

$$v_r^{\text{I}} + v_r^{\text{III}} = v_\theta^{\text{I}} + v_\theta^{\text{III}} = 0, \text{ for } r = a \text{ and also for } r = \infty \quad (29)$$

The artifice of introducing Stokes' solution as a part of the complete solution carries the advantage that the velocity  $V$  is eliminated. The particle velocity, therefore, will not enter explicitly into the solutions I and III. This fact justifies the use of equation 1', which implies that  $F^{\text{I}} + F^{\text{III}} = qE$  is independent of  $V$ .

<sup>20</sup> Debye, P., and Hückel, E., *Physik. Z.*, 25, 40 (1924).

Lamb<sup>21</sup> has shown that a general solution of the homogeneous equations for arbitrary boundary conditions can be represented by a sum of spherical harmonics, and it is sufficient, for the present purpose, to consider only the first terms of this series. It leads to

$$\begin{aligned} p^{\text{III}} &= A_0 + (A_1 r + B_1/r^2) \cos \theta \\ v_r^{\text{III}} &= (A_1 r^2/10\eta + B_1/\eta r + A_2 + B_2/r^3) \cos \theta \\ v^{\text{III}} &= (-A_1 r^2/5\eta - B_1/2\eta r - A_2 + B_2/2r^3) \sin \theta \end{aligned} \quad (30)$$

There remains therefore only the determination of a particular solution of the inhomogeneous equations. This particular solution is not restricted by boundary conditions, because 29 can always be satisfied by a proper choice of the constants  $A_i$  and  $B_i$  in 30.

The expressions on the right side of both equations 22' and 24' depend on the electric potential distribution  $\Phi$ . At a large distance from the particle both expressions will vanish, because  $\nabla^2 \Phi$  tends rapidly toward zero. The solution of the problem, therefore, requires only that the potential distribution be known accurately in the neighborhood of the micelle. It is this electrical problem which presents the greatest difficulties and which has not yet been solved in a rigorous manner. The theories of Hückel<sup>22</sup> and of Henry<sup>23</sup> assume that the electric field is given as a superposition of the field which exists around the charged particle when it is at rest, and of the field which the applied outer voltage would create when the particle was uncharged. This assumption, though plausible, is subject to criticism. It implies that the potential can be represented by

$$\Phi = \psi + \varphi \quad (31)$$

where the potential  $\psi$  is that of the electrical double layer and is supposed to possess spherical symmetry, i.e.,  $\frac{\partial \psi}{\partial \theta} = 0$ . Since the potential  $\varphi$  of the applied field  $X$  contains no space charges, it satisfies Laplace's equation  $\nabla^2 \varphi = 0$ , and equation 22' reduces to

$$\nabla^2 p = \frac{D}{4\pi} \frac{\partial \varphi}{\partial r} \frac{\partial (\nabla^2 \psi)}{\partial r} + \frac{D}{4\pi} \left\{ \frac{\partial \psi}{\partial r} \frac{\partial (\nabla^2 \psi)}{\partial r} + (\nabla^2 \psi)^2 \right\} \quad (32)$$

Its solution must be of the form  $p = p_0(r) + p_1(r, \theta)$ , where

$$\nabla^2 p_0 = \frac{D}{4\pi} \left\{ \frac{\partial \psi}{\partial r} \frac{\partial (\nabla^2 \psi)}{\partial r} + (\nabla^2 \psi)^2 \right\} \quad (33)$$

and

$$\nabla^2 p_1 = \frac{D}{4\pi} \frac{\partial \varphi}{\partial r} \frac{\partial (\nabla^2 \psi)}{\partial r} \quad (34)$$

<sup>21</sup> Lamb, H. *Hydrodynamics*, 6th edition, Cambridge University Press, 1932.

<sup>22</sup> Hückel, E. *Z. phys. Chem.*, **25**, 206 (1900).

<sup>23</sup> Henry, O. *Proc. Roy. Soc. London*, **A133**, 106 (1931).

Inasmuch as the pressure distribution,  $p_0$ , has spherical symmetry, it cannot give rise to a force on the particle, and therefore can be neglected in the electrophoretic problem. As was pointed out previously,  $p_0$  creates an increase of the hydrostatic pressure on the micelle. The second equation, 24', reduces to

$$\frac{\partial^2(rw)}{\partial r^2} + \frac{1}{r} \frac{\partial}{\partial \theta} \left[ \frac{1}{\sin \theta} \frac{\partial}{\partial \theta} (w \sin \theta) \right] = -\frac{D}{4\pi\eta} \frac{\partial \varphi}{\partial \theta} \frac{\partial(\nabla^2 \psi)}{\partial r} \quad (35)$$

and the theories of Hückel and of Henry deal primarily with the solutions of 34 and 35. The two theories differ in only one important respect. In the evaluation of the potential  $\varphi$  Debye and Hückel<sup>20</sup> ignore the presence of the particle. They assume that the applied voltage creates a field which is everywhere uniform. Hence they use

$$\begin{aligned} \varphi &= -Xr \cdot \cos \theta \\ E_r &= X \cos \theta \quad E_\theta = -X \sin \theta \end{aligned} \quad (36)$$

Henry, on the other hand, takes into account that the particle alters the field distribution. Henry postulates that

$$\varphi = -Xr \left( 1 + \lambda \frac{a^3}{r^3} \right) \cos \theta \quad (37)$$

This represents the potential distribution around an uncharged spherical particle when it is immersed in a conductive liquid across which a uniform field,  $X$ , is applied.  $\lambda = (\mu - \mu')/(2\mu + \mu')$ , where  $\mu$ ,  $\mu'$  are the specific conductivity of the solution and of the material of the particle, respectively. For dielectric particles,  $\mu' = 0$  and hence  $\lambda = \frac{1}{2}$ . The field at the surface of a dielectric particle is  $E_r = 0$ ,  $E_\theta = -\frac{3}{2}X \sin \theta$ , i.e., the tangential field component in Henry's theory is  $\frac{3}{2}$  times larger than in Hückel's. It is this circumstance which alters the factor  $4\pi$  in Smoluchowski's equation to  $6\pi$  in Hückel's theory.

If  $\mu = \mu'$ ,  $\lambda = 0$ . Since in this limiting case the two theories become identical, it is sufficient to give here Henry's solutions of the inhomogeneous equations. They are

$$\begin{aligned} p^I &= -\frac{DX}{4\pi} \cos \theta \left[ 3 \frac{\partial \psi}{\partial r} - 2\xi \right] \\ w^I &= -\frac{DX}{4\pi\eta} \xi \sin \theta \\ v_r^I &= \frac{DX}{6\pi\eta} \cos \theta \left[ \int^r \xi dr - \frac{1}{r^3} \int^r r^3 \xi dr \right] \\ v_\theta^I &= -\frac{DX}{6\pi\eta} \sin \theta \left[ \int^r \xi dr + \frac{1}{2r^3} \int^r \xi dr \right] \end{aligned} \quad (38)$$

where

$$\xi = \frac{\partial \psi}{\partial r} + \lambda a^3 r \int_{\infty}^r \frac{1}{r^4} \nabla^2 \psi dr \quad (39)$$

The lower limits of the integrals  $\int \xi dr$  and  $\int r^3 \xi dr$  are arbitrary parameters. The force component,  $F^I$ , is independent of these parameters and is found from 25 and 26 to have the value

$$F^I = -DXa^2 \left( \frac{\partial \psi}{\partial r} \right)_{r=a} = QX \quad (40)$$

because, by Gauss' theorem,  $-Da^2 \left( \frac{\partial \psi}{\partial r} \right)_{r=a} = Q$  is the charge of the particle, or the total charge of the ionic atmosphere.  $F^I$  represents the resultant force which the applied field exerts on the ions of the double layer. This force is transmitted to the particle and balances the electric force on the particle. As a consequence of this result the migration equation gives no direct information about the charge  $Q$  of the particle.

The last step of the calculation consists in satisfying the boundary conditions 29. This leads to a determination of the integration constants  $A_i$  and  $B_i$ , which are found to be  $A_0 = 0$ ,  $A_1 = 0$  and

$$A_2 = -\frac{2\eta B_1}{3a} = \frac{DX}{6\pi\eta} \int_{\infty}^a \xi dr$$

$$B_2 = \frac{DXa^3}{12\pi\eta} \int_{\infty}^a \xi dr + \frac{DX}{6\pi\eta} \int_{\infty}^a r^3 \xi dr$$

By introducing these values into 30 one finds from 25 and 26 the last component of the friction force

$$F^{II} = -DXa \int_{\infty}^a \xi dr \quad (41)$$

The total friction force on a spherical particle,  $F' = F^I + F^{II} + F^{III}$  is therefore

$$F' = 6\pi\eta aV + QX - DXa \int_{\infty}^a \xi dr \quad (42)$$

and comparison with 1' gives

$$q = Q - Da \int \xi dr$$

On the basis of Hückel's theory  $\lambda = 0$ ,  $\xi = \frac{\partial \psi}{\partial r}$ , whence  $q = Q - Da\psi_0$ , and from 5 and 3' it follows  $u' = (\psi'D - q/a)/6\pi\eta = 0$ . Hence Hückel's



theory leads to the same mobility equation, 4, which holds for a spherical particle in an insulating medium

$$u = \psi_0 D / 6\pi\eta.$$

Henry's theory, however, does not confirm this equation. Since  $F'$  balances the electric force  $QX$  on the particle, equation 42 furnishes  $u = V/X = \frac{D}{6\pi\eta} \int \xi dr$  and integration by parts of 39 leads to the following alternate expressions

$$u = \frac{D}{6\pi\eta} \left[ \psi_0 + \lambda a^3 \left( 3a^2 \int_{\infty}^a \frac{1}{r^5} \frac{\partial \psi}{\partial r} dr - 2 \int_{\infty}^a \frac{1}{r^3} \frac{\partial \psi}{\partial r} dr \right) \right] \quad (43a)$$

$$u = \frac{D}{6\pi\eta} \left[ \psi_0(1 + \lambda) + 3\lambda a^3 \left( 5a^2 \int_{\infty}^a \frac{\psi}{r^6} dr - 2 \int_{\infty}^a \frac{\psi}{r^4} dr \right) \right] \quad (43b)$$

According to this result, the electrophoretic migration speed is, in general, not only a function of the potential,  $\psi_0$ , of the particle, but depends also on the radius and the conductivity of the particle and on the course of the  $\psi(r)$  curve, *i.e.*, it depends on the structure of the double layer. The equation assumes a simple form if the potential can be represented by  $\psi = c/r^n$ , namely

$$u = \frac{D\psi_0}{6\pi\eta} [1 + \lambda n(n-1)/(n+5)(n+3)] \quad (44)$$

This equation is particularly well adapted to illustrate the significance of Henry's theory. It reduces to Hückel's equation 4 not only when  $\lambda = 0$ , but also when  $n = 1$ . Since the latter case corresponds to the potential of a particle in an insulating liquid, we are assured that the limiting case of a particle with a very thick double layer is given correctly by this theory. The other limiting case of a very thin double layer corresponds to a very steep potential gradient, *i.e.*, to a large value of  $n$ . If  $n$  is large, equation 44 furnishes  $u = \frac{D\psi_0}{6\pi\eta} (1 + \lambda)$ , and for a dielectric particle, for which  $\lambda = \frac{1}{2}$ ,

one obtains Smoluchowski's equation  $u = \frac{D\psi_0}{4\pi\eta}$ . This limiting law again satisfies our expectations, because if the extent of the double layer is small compared to the radius of the particle, the flow along its surface differs very little from that along a plane surface, for which Smoluchowski's equation has been shown to be valid.

For the intermediate case, where the radius of the particle and the thickness of the double layer are of the same order of magnitude, the potential variation cannot be represented by a law  $c/r^n$ . To a first approximation the potential is given by the Debye-Hückel law  $\psi = \psi_0 a e^{-\kappa(r-a)/r}$ , where  $\kappa$  is the reciprocal of the thickness of the double layer. As Henry has

shown, this potential distribution leads for dielectric particles to a migration equation of the form

$$u = \frac{D\psi_0}{4\pi\eta} f(\kappa a) \quad (45)$$

The function  $f(\kappa a)$  is obtained in the form of a power series from 43a or 43b

$$f(\kappa a) = 1 - 3/\kappa a + 25/(\kappa a)^2 - 220/(\kappa a)^3 + \dots \quad (46a)$$

or

$$f(\kappa a) = 1 + (\kappa a)^2/16 - 5(\kappa a)^3/48 - (\kappa a)^4/96 + (\kappa a)^5/96 - \frac{11}{96} e^{\kappa a} \int_a^{\infty} \frac{e^{-t}}{t} dt \quad (46b)$$

46a is applicable when  $(\kappa a) > 25$ ; 46b is convergent for all values of  $(\kappa a)$ , but the computation is practicable only when  $(\kappa a) < 5$ . This function is very closely approximated in its entire range by

$$f(\kappa a) = (16 + 7\kappa a + \kappa^2 a^2)/(6 + \kappa a)(4 + \kappa a) \quad (46c)$$

This approximation results from the fact that near the surface of the particle the Debye-Hückel potential can be approximated by  $c/r^n$ , where  $n = 1 + \kappa a$ . 46c is obtained by introducing this value of  $n$  into (44) and by using  $\lambda = \frac{1}{2}$ . The surprisingly close agreement between the results obtained by 46c and those derived from 46a or 46b is a verification of the statement that it is sufficient to know the potential distribution near the surface of the particle in order to arrive at a satisfactory equation for the electrophoretic mobility.

Figure 1 illustrates the course of the  $f(\kappa a)$  curve. For  $\kappa a > 300$ , Smoluchowski's equation is accurate to within 1 per cent. To within the same accuracy Hückel's equation becomes valid when  $\kappa a < 0.5$ . Although Henry's equation leads in this case to the same result as Hückel's theory, it should be noted that the two theories are based on different assumptions. Since Hückel's theory neglects the distortion of the field by the particle and thus employs a rather poor representation of the field distribution, which is especially poor in the range where the field is most effective, Henry's derivation should be considered preferable. On the other hand, it is doubtful whether Henry's theory is justified for any value of  $\lambda$  different from  $\frac{1}{2}$ . According to equation 37, the assumption  $\lambda \neq \frac{1}{2}$  implies that the electric current passes from the solution through the particle. This would necessitate the discharge of ions along the surface of the particle and this would lead to electrochemical reactions. Since no such phenomena have been observed, it is preferable to assume that even in the case of conducting micelles the electric field is tangential to the surface, as it is when  $\lambda = \frac{1}{2}$ . The assumption of a variable value of  $\lambda$  in Henry's theory should

be considered a mathematical artifice, which was introduced for the purpose of verifying Hückel's equation for the limiting case  $\lambda = 0$ .

Henry's theory is not a rigorous solution of the problem of electrophoretic migration. It is a first approximation, and the fact that it satisfies all limiting requirements shows that it must be a very good approximation. The objections and criticism of Henry's theory are directed against the assumption that the electric field around the particle is adequately represented by a superposition of the field in the double layer of a particle at rest and of the field as created by the electric current around an uncharged particle. This assumption, though it appears plausible, is rather arbitrary. Mooney<sup>24</sup> and Henry<sup>23</sup> have attempted to formulate a more

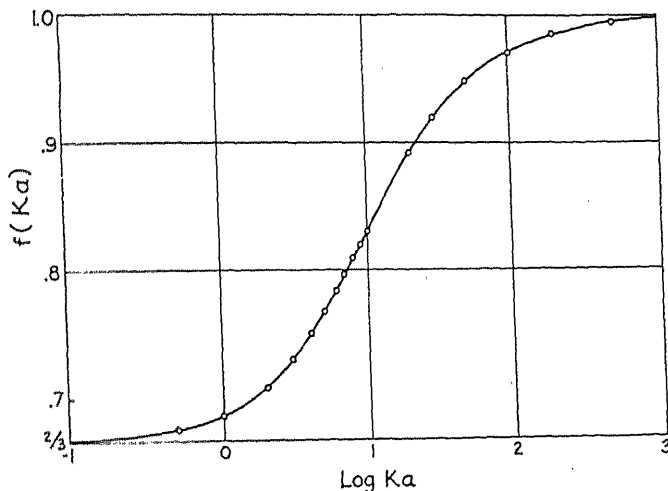


Figure 1.  $f(ka)$  vs.  $\log ka$  (from approximation eq. 46c)

accurate theory, but both find that it leads to differential equations which are too complicated to allow any hope of a complete solution. The difficulties become apparent when one realizes the complex interaction between hydrodynamic flow, electric current and electric charge distribution around the particle. The flow of the liquid and the electric current continually carry new ions to and from the environment of the micelle. The ions which constitute the ionic atmosphere are thus continually replaced by new ones. This exchange produces a distortion of the double layer because the ions which are approaching the micelle require a finite time before their distribution can adjust itself to the field distribution near the particle. Similarly, the ions which leave the double layer cannot instant-

<sup>24</sup> Mooney, M., *J. Physical Chem.*, 35, 331 (1931).

neously assume a random distribution. Hence the double layer will trail behind the moving particle; in front of the micelle the ionic atmosphere will be thinner than in the region behind the particle. The production of an asymmetric double layer around a moving particle is frequently referred to as the "relaxation effect" of the double layer.

An exact theory of this effect should take into account that the migration of the ions is due partly to the flow of the liquid, and partly to the applied voltage. Hence, the electric current distribution depends not only on the electric potential variations, but also on the hydrodynamic flow. Moreover, the current depends on the charge distribution in the double layer, because the latter determines the ionic concentrations in the neighborhood of the micelle. The combined effect of the flow and the charge distribution on the current results in an apparent change of the conductivity of the

TABLE 1. VALUES OF THE FUNCTION  $f(\kappa a) = \frac{16 + 7\kappa a + \kappa^2 a^2}{(6 + \kappa a)(4 + \kappa a)}$

Asymptotic behavior for

$$\text{small } \kappa a: f(\kappa a) = \frac{2}{3} \left[ 1 + \frac{\kappa a}{48} + \frac{7(\kappa a)^2}{576} - \frac{41(\kappa a)^3}{6912} + \dots \right]$$

$$\text{large } \kappa a: f(\kappa a) = 1 - \frac{3}{\kappa a} + \frac{22}{(\kappa a)^2} - \frac{148}{(\kappa a)^3} + \dots$$

$\kappa a$	$f(\kappa a)$	$\kappa a$	$f(\kappa a)$	$\kappa a$	$f(\kappa a)$
0	0.66667	7	0.79720	25	0.90767
0.1	0.6681	8	0.80952	30	0.91994
0.5	0.6752	9	0.82051	40	0.9367
1	0.68571	10	0.83036	50	0.9488
2	0.70833	12	0.84722	100	0.9702
3	0.73016	14	0.86111	200	0.9855
4	0.75000	16	0.87273	500	0.9941
5	0.76768	18	0.88257	1000	0.9970
6	0.78333	20	0.89103		

electrolyte in the neighborhood of the micelle. It gives rise to a surface conductivity. As Mooney<sup>24</sup> first pointed out, the surface conductivity will influence the electric field distribution. If, finally, we remember that on account of the relaxation effect the double layer in turn is altered by the current and, by virtue of the fundamental equations, the flow depends on the field and the charge distribution, we can arrive at a full appreciation of the complexity of the mathematical problem. To be sure, it is likely that various aspects of these intricate interactions may be disregarded on the ground that they would lead only to minor corrections, or that they might lead to deviations from the observed linear relationship between migration speed and applied field. Unfortunately, such a process of elimination has so far not been successful.

A simplified theory of the relaxation effect of the double layer around

ions was proposed by Debye and Hückel<sup>25</sup>. Their theory disregards the influence of the flow of the liquid and of the surface conductivity and considers only the asymmetry of the ionic atmosphere which is produced by a uniform current distribution. As a further simplification it is assumed that the asymmetry of the double layer does not alter either the flow or the friction force. Its effect consists in changing the electrical force on the particle. Paine<sup>26</sup> and Hermans<sup>27</sup> have adapted this theory to the case of colloidal particles, and the latter arrives at the conclusion that the distortion of the double layer will reduce the migration speed if the electrolyte is symmetric, but that in an asymmetric electrolyte an increase of the speed might eventually result. According to Hermans the relaxation effect plays an important role even if the double layer is much thinner than the radius of the particles and it may alter the migration velocity by as much as 50%. It seems rather doubtful whether these conclusions can be accepted. They are based on assumptions which appear far too greatly simplified. While a reduction of the migration-speed as the result of the relaxation effect cannot be questioned, it is not obvious that this effect can be adequately evaluated by considering only the electric forces between the asymmetric double layer and the charges of the particle.

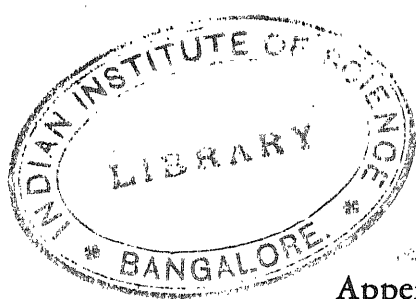
The present state of the theory of electrophoretic migration can be summarized as follows: If the thickness of the double layer is smaller than the radius of the particle the application of Smoluchowski's equation is justified for all particle shapes. In the other limiting case of small particles with a thick double layer, the migration formula approaches the form of Hückel's equation. For intermediate cases the factor  $C$  in the mobility equation  $u = \psi_0 D / C \eta$  will depend on the size, shape and charge of the micelle and on the constitution of the double layer. Although according to Henry's theory the value of  $C$  is between  $4\pi$  and  $6\pi$ , it is not impossible that on account of the relaxation effect  $C$  may in some cases exceed the value  $6\pi$ .

<sup>25</sup> Debye, P., and Hückel, E., *Physik. Z.*, 24, 305 (1923).

<sup>26</sup> Paine, H. H., *Proc. Cambridge Phil. Soc.*, 28, 83 (1932).

<sup>27</sup> Hermans, J. J., *Phil. Mag.*, (7) 25, 426 (1938); (7) 26, 650 (1938).





## Appendix

Dielectric Increments of Ellipsoidal Molecules, expressed as Functions of the Frequency of the Applied Field, and of the Axial Ratio of the Ellipsoid.

TABLE 1  
Axial Ratio  $a/b = 1$

Relative Frequency, $\nu/\nu_0$	Relative Dielectric Increment, $(D' - D_\infty)/(D_0 - D_\infty)$
10.0	0.010
8.0	.016
6.3	.025
5.0	.038
4.0	.049
3.2	.089
2.5	.138
2.0	.200
1.6	.281
1.25	.390
1.00	.500
0.80	.610
.63	.716
.50	.800
.40	.862
.32	.908
.25	.941
.20	.962
.16	.975
.125	.984
.100	.990
.080	.994
.063	.996
.050	.997
.040	.998
.032	.999
.025	1.000
.020	1.000
.016	1.000
.0125	1.000
.010	1.000

TABLE 2

Axial Ratio  $a/b = 2$ ;  $\nu_0/\nu_a = 1.505$ ;  $\nu_0/\nu_b = 1.050$

[illegible]

TABLE 3

Axial Ratio  $a/b = 4$ ;  $\nu_0/\nu_a = \tau_a/\tau_0 = 3.396$ ;  $\nu_0/\nu_b = \tau_b/\tau_0 = 1.190$ 

Increment Ratio, $\Delta D_a/\Delta D_b$ .....	0	0.1	0.25	0.5	1	2	4	10	$\infty$
Dipole Angle, $\theta^\circ$ ....	90.0	72.4	63.4	54.7	45.0	35.3	26.6	17.6	0
Relative Frequency, $\nu/\nu_0$	Relative Dielectric Increment, $(D' - D_\infty)/(D_0 - D_\infty)$								
10.0	0.007	0.006	0.006	0.005	0.004	0.003	0.002	0.002	0.001
8.0	.011	.010	.009	.008	.006	.004	.003	.002	.001
6.3	.018	.017	.015	.013	.010	.007	.005	.003	.002
5.0	.027	.025	.022	.019	.015	.012	.009	.006	.004
4.0	.042	.039	.035	.030	.024	.018	.013	.009	.006
3.2	.065	.060	.054	.046	.037	.028	.020	.014	.009
2.5	.101	.093	.084	.072	.058	.043	.031	.022	.014
2.0	.150	.138	.124	.107	.086	.064	.047	.033	.021
1.6	.216	.199	.179	.155	.124	.094	.070	.050	.033
1.25	.312	.288	.260	.226	.182	.139	.105	.077	.053
1.00	.414	.384	.347	.303	.247	.191	.147	.101	.080
0.80	.524	.487	.443	.389	.322	.254	.200	.156	.119
.63	.640	.598	.548	.486	.410	.333	.271	.221	.179
.50	.739	.695	.643	.578	.498	.418	.353	.301	.257
.40	.816	.774	.723	.661	.584	.507	.445	.394	.352
.32	.873	.835	.790	.734	.665	.596	.540	.495	.457
.25	.919	.888	.852	.807	.750	.694	.649	.613	.582
.20	.947	.923	.894	.859	.816	.772	.737	.708	.684
.16	.966	.943	.927	.901	.869	.837	.811	.790	.772
.125	.978	.966	.952	.935	.913	.891	.874	.860	.848
.100	.986	.978	.968	.956	.941	.926	.914	.904	.896
.080	.991	.986	.979	.971	.961	.951	.943	.936	.931
.063	.994	.991	.986	.981	.975	.969	.964	.959	.956
.050	.996	.994	.991	.988	.984	.980	.977	.974	.972
.040	.997	.996	.994	.992	.989	.986	.984	.982	.981
.032	.998	.997	.996	.995	.993	.991	.990	.989	.988
.025	.999	.999	.997	.997	.996	.995	.994	.994	.993
.020	1.000	1.000	.999	.998	.998	.997	.996	.995	.995
.016	1.000	1.000	.999	.999	.999	.998	.998	.997	.997
.0125	1.000	1.000	1.000	1.000	.999	.999	.998	.998	.998
.0100	1.000	1.000	1.000	1.000	1.000	.999	.999	.999	.999



TABLE 4

Axial Ratio  $a/b = 6$ ;  $\nu_0/\nu_a = \tau_a/\tau_0 = 6.062$ ;  $\nu_0/\nu_b = \tau_b/\tau_0 = 1.250$ 

Increment Ratio, $\Delta D_a/\Delta D_b$ .....	0	0.1	0.25	0.5	1	2	4	10	$\infty$
Dipole Angle, $\theta^\circ$ ....	90.0	72.4	63.4	54.7	45.0	35.3	26.6	17.6	0
Relative Frequency, $\nu/\nu_0$ .....	Relative Dielectric Increment, $(D' - D_\infty)/(D_0 - D_\infty)$								
10.0	0.006	0.005	0.005	0.004	0.003	0.002	0.001	0.001	0.0
8.0	.010	.009	.008	.007	.005	.003	.002	.001	0.0
6.3	.016	.015	.013	.011	.008	.006	.004	.002	.001
5.0	.025	.023	.020	.017	.013	.009	.006	.003	.001
4.0	.038	.035	.031	.026	.020	.014	.009	.005	.002
3.2	.059	.054	.048	.040	.032	.022	.014	.008	.003
2.5	.093	.085	.075	.063	.048	.034	.022	.012	.004
2.0	.136	.124	.110	.093	.071	.049	.032	.018	.006
1.6	.200	.183	.162	.137	.106	.074	.049	.028	.011
1.25	.291	.266	.236	.200	.154	.108	.072	.042	.017
1.00	.390	.357	.317	.269	.208	.147	.099	.059	.026
0.80	.500	.458	.408	.347	.270	.194	.133	.083	.041
.63	.624	.573	.512	.437	.344	.251	.176	.115	.064
.50	.719	.663	.595	.512	.408	.305	.222	.154	.098
.40	.800	.741	.669	.582	.473	.364	.277	.204	.146
.32	.862	.803	.731	.644	.536	.427	.340	.268	.209
.25	.912	.857	.790	.709	.608	.507	.426	.359	.304
.20	.941	.892	.834	.762	.673	.584	.512	.454	.405
.16	.962	.921	.873	.813	.738	.664	.604	.556	.515
.125	.976	.945	.908	.862	.806	.749	.703	.666	.635
.100	.984	.961	.934	.900	.858	.816	.782	.755	.732
.080	.990	.974	.954	.930	.900	.870	.846	.826	.810
.063	.994	.983	.970	.954	.934	.913	.897	.884	.873
.050	.996	.989	.980	.970	.956	.943	.933	.924	.917
.040	.997	.992	.987	.980	.971	.962	.955	.950	.945
.032	.998	.995	.991	.987	.981	.975	.971	.967	.964
.025	.999	.997	.995	.992	.988	.984	.981	.979	.977
.020	.999	.998	.996	.995	.993	.990	.989	.987	.986
.016	1.000	.999	.998	.997	.995	.993	.992	.991	.990
.0125	1.000	.999	.999	.998	.997	.996	.995	.995	.994
.0100	1.000	1.000	.999	.999	.998	.997	.997	.996	.996

TABLE 5

Axial Ratio  $a/b = 9$ ;  $\nu_0/\nu_a = \tau_a/\tau_0 = 11.58$ ;  $\nu_0/\nu_b = \tau_b/\tau_0 = 1.288$ 

Increment Ratio, $\Delta D_a/\Delta D_b$ .....	0	0.1	0.25	0.5	1	2	4	10	$\infty$
Dipole Angle, $\theta^\circ$ ....	90.0	72.4	63.4	54.7	45.0	35.3	26.6	17.6	0
Relative Frequency, $\nu/\nu_0$	Relative Dielectric Increment, $(D' - D_\infty)/(D_0 - D_\infty)$								
10.0	0.006	0.005	0.005	0.004	0.003	0.002	0.001	0.001	0.000
8.0	.009	.008	.007	.006	.005	.003	.002	.001	.000
6.3	.015	.014	.012	.010	.008	.005	.003	.002	.000
5.0	.024	.022	.019	.016	.012	.008	.005	.002	.000
4.0	.036	.033	.029	.024	.018	.012	.007	.004	.000
3.2	.056	.051	.045	.037	.028	.019	.011	.005	.000
2.5	.088	.080	.071	.059	.044	.030	.018	.009	.001
2.0	.130	.118	.104	.087	.066	.045	.028	.014	.002
1.6	.191	.174	.153	.128	.097	.066	.041	.020	.003
1.25	.278	.253	.223	.187	.142	.096	.060	.030	.005
1.00	.379	.344	.303	.255	.193	.131	.081	.041	.007
0.80	.485	.442	.390	.324	.248	.169	.106	.054	.011
.63	.604	.551	.487	.409	.312	.214	.136	.072	.019
.50	.707	.645	.571	.481	.368	.255	.165	.091	.029
.40	.790	.722	.641	.542	.418	.293	.194	.113	.045
.32	.855	.784	.698	.593	.462	.330	.225	.140	.068
.25	.907	.834	.747	.640	.506	.373	.266	.178	.106
.20	.938	.867	.782	.678	.543	.417	.313	.228	.157
.16	.960	.893	.813	.715	.593	.471	.373	.293	.226
.125	.975	.916	.845	.766	.649	.540	.453	.382	.323
.100	.983	.932	.872	.797	.704	.615	.538	.476	.426
.080	.990	.949	.900	.838	.764	.689	.628	.579	.538
.063	.994	.963	.926	.880	.823	.766	.720	.683	.652
.050	.996	.974	.947	.914	.874	.833	.800	.773	.751
.040	.997	.981	.963	.940	.911	.882	.859	.841	.825
.032	.998	.987	.974	.959	.939	.919	.904	.891	.880
.025	.999	.992	.984	.973	.962	.948	.937	.929	.922
.020	.999	.994	.989	.982	.974	.966	.953	.954	.949
.016	1.000	.997	.994	.989	.984	.978	.974	.971	.968
.0125	1.000	.998	.996	.993	.990	.986	.983	.982	.979
.0100	1.000	.999	.997	.999	.994	.991	.990	.988	.987

TABLE 6

Axial Ratio  $a/b = 12$ ;  $\nu_0/\nu_a = \tau_a/\tau_0 = 17.93$ ;  $\nu_0/\nu_b = \tau_b/\tau_0 = 1.305$ 

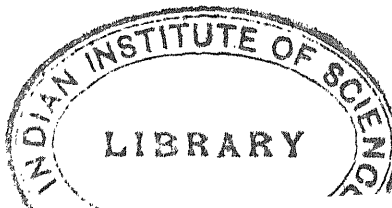
Increment Ratio, $\Delta D_a/\Delta D_b, \dots$	0	0.1	0.25	0.5	1	2	4	10	$\infty$
Dipole Angle, $\theta^\circ, \dots$	90.0	72.4	63.4	54.7	45.0	35.3	26.6	17.6	0
Relative Frequency, $\nu/\nu_0$	Relative Dielectric Increment, $(D' - D_\infty)/(D_0 - D_\infty)$								
10.0	0.006	0.006	0.005	0.004	0.003	0.002	0.001	0.001	0.000
8.0	.009	.008	.007	.006	.005	.003	.002	.001	.000
6.3	.015	.014	.012	.010	.008	.005	.003	.001	.000
5.0	.023	.021	.018	.015	.012	.008	.005	.002	.000
4.0	.035	.032	.028	.023	.018	.012	.007	.003	.000
3.2	.055	.050	.044	.037	.028	.018	.011	.005	.000
2.5	.086	.078	.069	.058	.043	.029	.018	.009	.001
2.0	.128	.116	.103	.086	.064	.043	.026	.013	.001
1.6	.186	.169	.149	.124	.093	.063	.038	.018	.001
1.25	.272	.247	.218	.182	.137	.092	.056	.026	.002
1.00	.370	.337	.297	.248	.186	.126	.076	.036	.003
0.80	.478	.435	.383	.320	.242	.163	.100	.048	.005
.63	.597	.543	.479	.401	.303	.204	.126	.062	.008
.50	.701	.638	.563	.471	.356	.246	.150	.075	.012
.40	.786	.716	.633	.530	.402	.275	.172	.089	.019
.32	.851	.776	.687	.577	.440	.303	.193	.104	.029
.25	.905	.827	.733	.619	.476	.333	.219	.125	.047
.20	.936	.857	.763	.648	.504	.360	.245	.151	.072
.16	.962	.884	.791	.677	.535	.393	.279	.186	.108
.125	.974	.900	.812	.705	.570	.435	.328	.239	.166
.100	.983	.915	.834	.735	.610	.486	.387	.306	.238
.080	.989	.929	.857	.769	.659	.548	.460	.388	.328
.063	.994	.944	.883	.809	.716	.624	.550	.489	.439
.050	.996	.956	.908	.850	.776	.703	.645	.597	.557
.040	.997	.966	.930	.885	.828	.772	.727	.691	.660
.032	.998	.976	.949	.916	.876	.835	.802	.775	.753
.025	.999	.984	.966	.944	.916	.888	.866	.848	.833
.020	.999	.989	.976	.961	.942	.923	.908	.895	.885
.016	1.000	.993	.985	.974	.962	.949	.938	.930	.923
.0125	1.000	.996	.990	.984	.976	.968	.962	.956	.952
.0100	1.000	.998	.994	.990	.984	.979	.975	.972	.969
.0080	1.000	.998	.996	.993	.990	.986	.983	.981	.979
.0063	1.000	.999	.998	.996	.994	.992	.990	.989	.988
.0050	1.000	.999	.998	.997	.996	.995	.994	.993	.992

TABLE 7. AXIAL RATIO  $a/b = 16$ ;  $\nu_0/\nu_a = \tau_a/\tau_0 = 28.78$ ;  $\nu_0/\nu_b = \tau_b/\tau_0 = 1.316$ 

Increment Ratio, $\Delta D_a/\Delta D_b$ .....	0	0.1	0.25	0.5	1	2	4	10	$\infty$
Dipole Angle, $\theta^\circ$ ..	90.0	72.4	63.4	54.7	45.0	35.3	26.6	17.6	0
Relative Frequency, $\nu/\nu_0$	Relative Dielectric Increment, $(D' - D_\infty)/(D_0 - D_\infty)$								
10.0	0.005	0.005	0.004	0.003	0.002	0.002	0.001	0.000	0.000
8.0	.009	.008	.007	.006	.004	.003	.002	.001	.000
6.3	.015	.014	.012	.010	.008	.005	.003	.001	.000
5.0	.023	.021	.018	.015	.012	.008	.005	.002	.000
4.0	.035	.032	.028	.023	.018	.012	.007	.003	.000
3.2	.054	.049	.043	.036	.027	.018	.011	.005	.000
2.5	.084	.076	.067	.056	.042	.028	.017	.008	.000
2.0	.126	.115	.101	.084	.063	.042	.025	.011	.000
1.6	.184	.167	.147	.123	.092	.062	.038	.018	.001
1.25	.269	.245	.215	.180	.135	.091	.055	.025	.001
1.00	.366	.333	.293	.244	.184	.123	.074	.034	.001
0.80	.475	.432	.380	.317	.238	.160	.101	.045	.002
0.63	.594	.540	.476	.396	.298	.200	.121	.057	.003
.50	.697	.634	.559	.466	.351	.236	.143	.068	.005
.40	.782	.712	.626	.524	.395	.266	.163	.078	.008
.32	.849	.773	.672	.570	.430	.291	.179	.088	.012
.25	.903	.823	.726	.608	.461	.314	.196	.099	.019
.20	.936	.854	.755	.634	.482	.331	.210	.111	.029
.16	.958	.875	.775	.654	.502	.349	.228	.128	.045
.125	.974	.892	.794	.673	.523	.373	.252	.154	.072
.100	.982	.903	.807	.690	.544	.399	.282	.187	.107
.080	.989	.913	.823	.712	.574	.436	.325	.234	.159
.063	.994	.925	.842	.741	.614	.487	.386	.303	.234
.050	.996	.935	.862	.773	.662	.550	.461	.388	.327
.040	.997	.945	.884	.808	.714	.619	.543	.482	.430
.032	.998	.956	.907	.846	.770	.693	.632	.582	.541
.025	.999	.968	.931	.885	.828	.772	.726	.689	.658
.020	1.000	.977	.951	.918	.876	.835	.802	.775	.753
.016	1.000	.984	.965	.942	.913	.884	.861	.842	.826
.0125	1.000	.990	.977	.962	.942	.923	.908	.895	.885
.0100	1.000	.993	.984	.974	.961	.948	.938	.929	.922
.0080	1.000	.995	.990	.983	.975	.967	.960	.954	.950
.0063	1.000	.997	.994	.990	.984	.979	.975	.972	.969
.0050	1.000	.998	.996	.993	.990	.986	.983	.981	.979
.0040	1.000	.999	.997	.996	.993	.991	.990	.988	.987
.0032	1.000	.999	.998	.997	.996	.995	.994	.993	.992
.0025	1.000	1.000	.999	.998	.997	.997	.996	.995	.995
.0020	1.000	1.000	.999	.999	.998	.997	.997	.996	.996
.0016	1.000	1.000	.999	.999	.998	.998	.998	.997	.997
.00125	1.000	1.000	1.000	.999	.999	.999	.998	.998	.998
.00100	1.000	1.000	1.000	1.000	1.000	1.000	.999	.999	.999

TABLE 8. AXIAL RATIO  $a/b = 25$ ;  $\nu_0/\nu_a = \tau_a/\tau_0 = 61.07$ ;  $\nu_0/\nu_b = \tau_b/\tau_0 = 1.325$ 

Increment Ratio, $\Delta D_a/\Delta D_b$ .....	0	0.1	0.25	0.5	1	2	4	10	$\infty$
Dipole Angle, $\theta^\circ$ ..	90.0	72.4	63.4	54.7	45.0	35.3	26.6	17.6	0
Relative Frequency, $\nu/\nu_0$	Relative Dielectric Increment, $(D' - D_\infty)/(D_0 - D_\infty)$								
10.0	0.006	0.005	0.005	0.004	0.003	0.002	0.001	0.001	0.000
8.0	.009	.008	.007	.006	.005	.003	.002	.001	.000
6.3	.014	.013	.011	.009	.007	.005	.003	.001	.000
5.0	.022	.020	.018	.015	.011	.007	.004	.002	.000
4.0	.034	.031	.027	.023	.017	.011	.007	.003	.000
3.2	.053	.048	.042	.035	.026	.018	.011	.005	.000
2.5	.084	.076	.067	.056	.042	.026	.017	.008	.000
2.0	.125	.114	.100	.083	.062	.042	.025	.011	.000
1.6	.182	.165	.146	.121	.091	.061	.036	.017	.000
1.25	.265	.241	.212	.177	.132	.088	.053	.024	.000
1.00	.363	.330	.291	.242	.182	.121	.073	.033	.000
0.80	.470	.427	.376	.313	.235	.157	.094	.043	.000
.63	.590	.536	.474	.394	.296	.197	.119	.055	.001
.50	.695	.632	.556	.464	.348	.232	.140	.064	.001
.40	.780	.709	.624	.521	.391	.261	.158	.073	.002
.32	.847	.770	.678	.566	.425	.284	.172	.080	.003
.25	.902	.820	.722	.603	.453	.303	.184	.086	.004
.20	.935	.850	.749	.625	.470	.316	.192	.090	.006
.16	.956	.870	.767	.641	.483	.325	.199	.096	.010
.125	.974	.887	.783	.655	.496	.336	.208	.104	.017
.100	.982	.895	.791	.663	.504	.345	.217	.113	.026
.080	.989	.903	.799	.673	.514	.356	.230	.126	.040
.063	.994	.909	.808	.684	.528	.373	.249	.148	.063
.050	.996	.914	.816	.696	.546	.396	.276	.178	.096
.040	.997	.919	.826	.713	.570	.428	.315	.222	.144
.032	.998	.926	.840	.734	.603	.471	.365	.279	.207
.025	.999	.935	.859	.766	.650	.533	.440	.364	.300
.020	1.000	.946	.880	.801	.700	.601	.522	.456	.402
.016	1.000	.956	.902	.837	.756	.674	.609	.555	.511
.0125	1.000	.966	.926	.877	.816	.755	.706	.665	.632
.0100	1.000	.975	.946	.909	.864	.819	.782	.753	.728
.0080	1.000	.982	.961	.936	.904	.871	.846	.825	.807
.0063	1.000	.988	.974	.957	.936	.914	.897	.883	.871
.0050	1.000	.992	.983	.972	.958	.943	.932	.923	.915
.0040	1.000	.995	.989	.981	.972	.963	.955	.949	.944
.0032	1.000	.997	.993	.988	.982	.978	.971	.967	.964
.0025	1.000	.998	.995	.992	.988	.985	.982	.979	.977
.0020	1.000	.999	.997	.995	.992	.990	.988	.986	.985
.0016	1.000	.999	.998	.997	.996	.994	.993	.992	.991
.00125	1.000	.999	.999	.998	.997	.996	.995	.994	.994
.00100	1.000	1.000	.999	.998	.998	.997	.996	.995	.995



## Table of Important Symbols

$a$	= activity.
$a_k$	= activity of the $k$ 'th component.
$a$	= cohesion term in the van der Waals equation (Chapter 3).
$a$	= length of semi-axis of revolution in an ellipsoidal molecule (Chapters 18, 19, 21, 22).
$a'$	= distance defined by equation 168, Chapter 3.
$a_{ij}$	= mean ionic diameter (Chapter 3).
$A = E - TS$	= work content of system.
$\bar{A}$	= surface area (Chapters 3 and 8).
$b$	= length of equatorial semi-axis of an ellipsoidal molecule (Chapters 18, 19, 21, 22).
$b$	= radius of an ion or dipolar ion.
$b$	= volume term in the van der Waals equation (Chapter 3).
$b_{ij}$	= coefficient in heat content equation for solution (Chapter 3).
$B_{ij}$	= coefficient in expression for activity coefficient in solution (Chapter 3).
$c$	= concentration in molecules $\text{cm}^{-3}$ .
$c_{ij}$	= coefficient in heat content equation for solution (Chapter 3).
$C$	= concentration in moles per liter.
$C_{ij}$	= coefficient in expression for activity coefficient in solution (Chapter 3).
$d$	= density (Chapter 7).
$d$	= distance = number of carbon atoms separating the dissociating group from the substituent (Chapter 4).
$d_{ijk}$	= coefficient in heat content equation for solution (Chapter 3).
$\partial$	= partial derivative.
$D$	= translational diffusion constant (especially in Chapters 18 and 19).
$D_{20}$	= diffusion constant reduced to standard conditions (water at $20^\circ\text{C}$ ). (Chapter 19).
$D$	= dielectric constant.

$D_0$	= dielectric constant in the standard state, or of the pure solvent, water.
$D_0$	= dielectric constant at low frequency (Chapter 22).
$D^0$	= dielectric constant of solvent (Chapter 22).
$D_\infty$	= dielectric constant at frequency much higher than critical frequency (Chapter 22).
$D_E$	= effective dielectric constant (Chapters 5 and 12).
$\Delta D$	= dielectric increment (Chapter 22).
$D_{ijh}$	= coefficient in expression for activity coefficient in solution (Chapter 3).
$e$	= base of natural logarithms.
$e$	= electrical charge.
$e_{ijh}$	= coefficient in heat content equation for solution (Chapter 3).
$E$	= electric field intensity (Chapter 25).
$E$	= electrostriction (Chapters 7 and 16).
$E$	= energy of system (Chapter 3).
$E_{ijh}$	= coefficient in expression for activity coefficient in solution (Chapter 3).
$EMF$	= electromotive force.
$\xi$	= electrical potential (Chapter 3).
$f$	= activity coefficient (in Chapter 9, pp. 196-198).
$f$	= restoring force (Chapter 2).
$f$	= frictional resistance to a molecule moving through a liquid with unit velocity.
$f_0$	= frictional resistance to an unhydrated spherical molecule moving through a liquid with unit velocity.
$F$	= force.
$F$	= force constant of a chemical bond (Chapter 2).
$F$	= frictional resistance to one mole of solute molecules moving through a liquid with unit velocity (Chapters 18 and 19).
$G = E + PV - TS$	= free energy of system (Chapters 3 and 7).
$G_k^0$	= free energy of $k$ 'th component in standard state (Chapter 3).
$F$	= value of faraday in coulombs.
$c$	= concentration of solute in grams per liter.
$\gamma$	= gravitational potential (Chapter 3).
$\gamma$	= any extensive property of the solution (Chapters 3 and 7).

$G_1$	= corresponding molal property of the solvent (Chapter 7).
$h$	= coefficient in equation relating dielectric constant and polarization (Chapters 6 and 22).
$h$	= number of protons dissociated from a protein molecule, $h$ being taken as zero in the state of maximum binding (Chapter 20).
$H$	= heat content of system.
$(H^+)$	= hydrogen ion activity.
$(H^+)$	= hydrogen ion activity at isoelectric point.
$\Delta H_f/n$	= heat of fusion per mole (Chapter 8).
$I/2$	= ionic strength, when ionic concentrations are expressed as molalities = $\frac{1}{2}\sum m_i z_i^2$ .
$J$	= $a/b$ ; axial ratio of an ellipsoid (Chapter 21).
$k$	= Boltzmann's constant = $R/N = 1.380 \times 10^{-16}$ erg deg <sup>-1</sup> .
$k$	= rate constant (Chapter 3, p. 73).
$k$	= with subscript 1, 2, 3, etc., or subscript 11, 12, etc. = dissociation constant of an individual acidic group in a protein (Chapter 20).
$k_\infty$	= dissociation constant of an individual acidic group, in a hypothetical medium of infinite dielectric constant (Chapter 20).
$K$	= equilibrium constant.
$K'$	= equilibrium constant in terms of weight molalities (Chapter 3).
$K_2$	= constant depending on the nature of the polar group (Chapter 9).
$K_R, K'_R$	= interaction constant of ion-dipolar ion due to Coulomb forces (salting-in coefficient).
$K_R^*$	= salting-in constant for dipole-dipole interaction (Chapter 10).
$K_s, K_s$	= salting-out constant for ion-dipole interaction.
$K_s^*$	= salting-out constant for dipole-dipole interaction (Chapter 10).
$K_z$	= activity ratio of dipolar ions to uncharged molecules (Chapter 4).
$\ln$	= logarithm to base $e$ .
$\log$	= logarithm to base 10, common logarithm.
$L$	= distance between plates of electrophoretic apparatus (Chapter 25).
$L_1, L_2$	= association constants for amino acids in formaldehyde solution (Chapter 5).



	= molality (moles per kilogram of solvent).
	= moles of $i$ 'th component per kilogram of solvent.
	= molality so low that deviation from ideal solution laws may be neglected (Chapter 3).
	= mean electric moment per molecule (Chapter 6, p. 143)
$f$	= molecular weight.
$f$	= total molality of all solutes, $\sum_i m_i$ (Chapter 3, equation 29, p. 26).
	= refractive index.
	= number of moles of $i$ 'th species in system.
	= number of ions in spherical shell of unit thickness (Chapter 3).
	= Avogadro's number = $6.022 \times 10^{23}$ .
	= mole fraction.
	= mole fraction of $k$ 'th component.
	= electric polarization per unit volume.
	= electric polarization of solute (Chapter 6).
	= polarization of solvent (Chapter 6).
	= pressure (Chapter 25).
	= partial pressure of second component (Chapter 3).
$G$	= titration constant (see Chapters 4 and 20).
$H$	= $-\log (H^+)$ ; see definition in Chapter 3, p. 43, equation 69.
$I$ (or $pH_I$ )	= $pH$ of isoelectric point.
$M$	= $pH$ of maximum charge (Chapter 4).
$K$	= negative logarithm of dissociation constant.
$K_{max}$	= value of $pK$ at temperature of maximum ionization.
	= molar polarization (Chapter 6).
	= pressure (Chapter 3).
$**$	= pressure so low that deviations from perfect gas laws are negligible.
	= area of surface across which diffusion takes place (p. 397).
	= average total charge (Chapter 4, p. 94).
	= charge on a particle (Chapter 25).
	= apparent heat of ionization in kcal/mole (Chapter 20).
	= distance.
	= equilibrium distance between two atoms (Chapter 2).

$R$	= dipole distance = $\mu/\epsilon$ .
$R$	= gas constant (Chapter 3).
$s$	= specific heat of solution (Chapter 7).
$s$	= sedimentation constant (Chapter 19, p. 422, equation 9).
$s_{20}$	= sedimentation constant reduced to standard conditions (water at 20°C) (See Chapter 19, p. 422, equation 9a).
$S$	= entropy of system (Chapter 3).
$S$	= mass of solute diffusing or sedimenting (Chapters 18 and 19).
$S$	= Svedberg unit of sedimentation constant = $10^{-13}$ sec (Chapter 19).
$S$	= solubility.
$S_n$	= solubility of neutral (isoelectric) protein molecules (Chapter 24).
$t$	= centigrade temperature.
$t$	= time.
$t_i$	= transference number of $i$ 'th species.
$T$	= temperature in °K.
$T_f$	= temperature of fusion (Chapter 8).
$T_{\max}$	= temperature of maximum ionization (Chapters 4 and 5).
$u$	= mobility.
$v$	= velocity of electrophoretic migration (Chapter 25).
$\bar{v}$	= partial specific volume (Chapter 22).
$v_x, v_y, v_z$	= components of electrophoretic velocity, $v$ (Chapter 25).
$V$	= migration velocity (Chapter 25)
$V$	= partial specific volume (Chapters 18 and 19).
$V$	= volume of system (Chapter 3).
$\bar{V}_k$	= partial molal volume of $k$ 'th component (Chapter 3).
$V_1$	= volume of water.
$w$	= hydration of protein, expressed as gram solvent bound per gram protein.
$w$	= velocity vector (Chapter 25, equation 23).
$w_1$	= one thousandth of molecular weight of first (solvent) component.
$W$	= mass of system.
$W_e$	= electrical work.
$W_i$	= molecular weight of $i$ 'th component (Chapter 3).

$x$	= space coördinate (Chapter 3).
$x_i$	= solute mole fraction of $i$ 'th species $m_i/M$ (Chapter 3).
$X$	= electric intensity of applied field (Chapter 25).
$y$	= space coördinate.
$Y$	= quantity defined by equation 131, Chapter 3 p. 57.
$z$	= space coördinate.
$z_i$	= valence of $i$ 'th ionic species.
$Z$	= net charge in proton units (Chapter 20).
$\bar{Z}$	= mean net charge in proton units (Chapter 20).
$\alpha$	= $(\beta/\theta)$ ratio of velocity gradient to rotary diffusion constant (Chapter 21).
$\alpha_0$	= electric polarizability (Chapter 6).
$\alpha_q$	= fraction of total number of groups of class $q$ which have dissociated protons (Chapters 5 and 20).
$\beta$	= compressibility (Chapters 3 and 7).
$\beta$ or $\beta'$	= extrapolated intercept of salting-out curve (Chapters 11 and 24).
$\beta$	= velocity gradient (Chapter 21).
$\beta_{ij}$	= coefficient in expression for work content of gas (Chapter 3).
$\Gamma$	= ionicity (double ionic strength, in moles per liter), $\sum_i C_i z_i^2$ .
$\Gamma_k$	= surface concentration of $k$ 'th component.
$\gamma$	= activity coefficient.
$\gamma_e$	= electrostatic contribution to the activity coefficient (Chapter 4).
$\delta$	= generalized measure of particle size (Chapter 3).
$\delta$	= molar dielectric increment.
$\delta_{ijk}$	= coefficient in expression for work content of gas (Chapter 3).
$\epsilon$	= proton charge = $4.803 \times 10^{-10}$ esu.
$\zeta$	= coefficient of frictional resistance to rotary molecular motion (especially in Chapter 21).
$\eta$	= viscosity.
$\eta_0$	= viscosity of solvent.
$\theta$	= rotary diffusion constant (Chapters 12 and 21).
$\theta$	= angular coördinate.
$\kappa$	= parameter of Debye-Hückel theory, defined by equation 108, p. 54, Chapter 3.
$\kappa$	= specific conductivity (Chapter 22).

$\lambda$	= a function of specific conductivity, defined on p. 637 (Chapter 25).
$\mu$	= dipole moment.
$\mu_k$	= dipole moment of the $k$ 'th component.
$\bar{\mu}$	= an electric moment defined in Chapter 12, p. 294.
$\tilde{\mu}$	= an electric moment defined in Chapter 12, p. 294.
$\mu_i$	= molal chemical potential of $i$ 'th species.
$\mu_0$	= chemical potential of $k$ 'th component in standard state.
$\nu$	= viscosity increment of solute, defined by equation 31, p. 515 (Chapter 21).
$\nu$	= frequency (Chapter 22).
$\nu_c$	= critical frequency.
$\nu_i$	= number of moles of $i$ 'th species produced (Chapter 3).
$\pi$	= osmotic pressure (Chapter 17, et seq.).
$\rho$	= density (Chapter 3).
$\rho$	= electrical density (Chapter 3, p. 53).
$\sigma$	= surface density of charge (Chapter 25).
$\sigma$	= surface tension.
$\Sigma_i$	= summation over all values of $i$ .
$\tau$	= relaxation time (Chapters 21 and 22).
$\Phi$	= potential function for applied electric field (Chapter 25).
$\bar{v}$	= volume fraction of system occupied by solute (Chapter 21).
$\Phi_{s2}$	= apparent molal volume (Chapter 3, p. 61).
$\Phi(C_p)$	= apparent molal heat capacity (Chapter 7).
$\Phi_{cp2}$	= relative apparent molal heat capacity (Chapter 3, p. 61).
$\Phi_{H2}$	= relative apparent molal heat content (Chapter 3, p. 61).
$\Phi(G)$	= apparent molal property $G$ (Chapter 7).
$\varphi$	= apparent molal volume of solute (in Chapter 12, pp. 295-296).
$\varphi$	= osmotic coefficient (see Chapter 3, equations 29 and 32).
$\phi$	= angular coördinate (Chapter 3).
$\phi$	= force acting on one gram mole of solute (Chapter 19).
$\psi$	= osmotic coefficient (Chapter 10).

$\chi$	= extinction angle (Chapter 21).
$\psi$	= electrical potential.
$\psi$	= angle of isocline (Chapter 21).
$\Omega$	= angular velocity (Chapter 21).
$\omega$	= angular velocity (Chapters 19 and 21).
$\omega$	= frequency (Chapter 12, p. 297).
$\nabla^2$	= Laplacian operator (Chapters 3 and 12).

510-6

71-90

7247



## Subject Index

The Tabular Index of Data for Certain Important Amino Acids and Peptides (pp. 680-683) and the Tabular Index of Data for Certain Important Proteins (pp. 684-686) form an essential supplement to this subject index. Amino acids and peptides listed in the tabular index are marked with an asterisk (\*) in this subject index, and in the tabular index will be found page references for certain important data concerning them. The tabular index for data for proteins lists certain important data for a selected group of proteins, which are marked with a dagger (†) in this subject index.

### A

- Acetamide, effect of, on sedimentation and diffusion of hemoglobin, 436  
partial specific volume of, 376  
solubility of, in water and organic solvents, 206
- Acetanilide, solubility of, in ethanol-water mixtures, 215
- Acetic acid, apparent molal heat capacity of, 169  
heats of ionization of, 117  
 $pK$  values of, in ethanol-water, 109  
in water, 109, 117
- Acetnaphthalide, solubility of, in ethanol-water mixtures, 215
- Acetone, solubility of amino acids, peptides and related compounds in, 201, 202, 208, 209, 212
- Acetyl glycine,  $pK'$  values of, 114, 135
- Acetylhistidine, apparent molal volumes and electrostriction of, 159
- Acid-base equilibria, determination of nature of ionizing groups in proteins by study of, 447-9  
effect of charged substituents on, 117-20, 287-94  
effect of dipolar substituents on, 120-4  
effect of ionic strength and dielectric constant on, 67-72, 105-15, 468-77  
fundamental concepts of acids and bases, 75-77  
in protein solutions, 444 ff.  
relation of, to solubility, 590-9
- see also:* Heat of ionization; Titration curves of proteins; Dissociation constants
- Acid combining capacity of proteins, values of, relation to content of basic amino acids, 354, 446, 478-505
- Acidity, definition of, 75-77
- Acids and bases, 75-77  
Brönsted's definition of, 75  
theory of G. N. Lewis concerning, 75  
*see also:* Acid-base equilibria
- Activity (thermodynamic), of protein in the crystalline phase, 579, 584-5  
*see also:* Activity coefficients
- Activity coefficients, definition of, 26  
of amino acids, peptides and related compounds, in aqueous solutions, 196-8, 217 ff.  
in electrolyte solutions, 236-75, 277-87  
in solutions of other dipolar ions, 217-35  
in water and organic solvents, 196-216  
of ions, in amino acid solutions, 251-65  
in electrolyte solutions, 59-60, 107-11, 468-77  
in protein solutions, 394-5, 619-22  
of one dipolar ion in the presence of others, 217-35, 394-5, 617-22  
of peptides, in solutions of electrolytes, 272-5  
in solutions of other dipolar ions, 226-35

- of proteins, deduced from EMF measurements, 468-77, 621-3
  - deduced from osmotic pressure studies, 394, 621
  - deduced from solubility measurements, 586-621
  - in the presence of salts, 584-5, 602-22
  - significance of, for titration curves, 468-77
- of salts, in solutions of dipolar ions, 254-65, 619-22
  - relation of, to dipole moments, in dipole-dipole interactions, 228-35
  - in ion-dipole interactions, 272-87
  - to osmotic coefficients, in a two-component system, 34
- $\alpha$ -Alanine,\* interatomic distances in, 320
- $\beta$ -Alanine\*
  - $\alpha$ -Alanine hydantoic acid, apparent molal volume of, 372
  - solubility of, in water and organic solvents, 201, 206, 208, 209
- $\beta$ -Alanine hydantoic acid, apparent molal volume of, 372
  - solubility of, in water and in ethanol, 201, 206
- Alanylalanine, activity coefficient of, 198, 228
  - $pK'$  values of, in water, from cells with liquid junction, 84
- Alanyldiglycine,  $pK'$  values of, in water, from cells with liquid junction, 84
  - relaxation time of, 566
- Alanylglycine, activity coefficient of, 198, 228
  - apparent molal volume and electrostriction of, 159
  - dielectric increment of, 146
  - $pK'$  values of, in water, from cells with liquid junction, 84
  - relaxation time of, 566
- Alanylleucylglycine, relaxation time of, 566
- Alanylproline,  $pK'$  values of, in water, from cells with liquid junction, 84
- Alcohols, activity coefficients of, in aqueous solution, 182
  - apparent molal heat capacity of, 169
  - see also: Ethanol, Butanol, etc.
- Alicyclic amino acids (aminocyclohexane carboxylic acids), melting points,  $pK'$  values,  $K_s$  values, solubility values of, 128
- Amandin,† molecular weight of, from osmotic pressure in urea solution, 390-1
  - viscosity increment of, 520
- Amino acid derivatives, melting points of, 185
- Amino acid esters, dipole moments of, 148
  - dissociation constants of, 97-9, 101, 102, 112-5
  - interconversion of, with isomeric betaines, 75 (footnote)
- Amino acid residues, calculation of specific volumes of, 372, 375
  - content of, in proteins, 368
  - volume fractions of, in proteins, 375
- Amino acid salts, apparent molal volume of, 163
- Amino acids, analysis of, by isotope dilution method, 365
  - by other methods, 356-9
  - by solubility of amino acid derivatives, 363-5
- apparent molal heat capacities of, 165-73
- apparent molal volumes of, 157-65
  - classes of, in proteins, 344-6
  - content of, in proteins, 348-69
  - dielectric constants of, 140-54
  - diffusion constants of, 410-2
  - dissociation constants of, from EMF measurements, on cells with liquid junction, 83-90
  - on cells without liquid junction, 78-83
  - heats of ionization of, 81-3, 88-90
  - heats of solution of, 192, 194
  - infrared spectra of, 19
  - isoelectric points of, 90
  - isolation of, from protein hydrolysates, 339
  - melting points of, 185, 362
  - partial specific volume of, 376
  - Raman spectra of, 11-9
  - relaxation times of, 565-7
  - solubility of, in electrolyte solutions, 236-75, 277-87
    - in organic solvents, 198-216
    - in solutions of other dipolar ions, 217-35
    - in water at various temperatures, 187-95, 199-202
  - see also: Activity coefficients

\* See Tabular Index of Amino Acids and Peptides, pp. 680-682.

† See Tabular Index of Proteins, p. 685.

- Amino acids, steric configuration of, 308  
 surface tension of solutions of, 174-6
- Aminobenzoic acids,\* effect of resonance  
 on dissociation constants of, 127-9  
 equilibrium between dipolar ions and  
 uncharged molecules in, 98-9,  
 127-9, 160
- $\alpha$ -Amino-isobutyric acid\*  
 $\alpha$ -Amino-*n*-butyric acid\*  
 $\beta$ -Aminobutyric acid\*  
 $\gamma$ -Aminobutyric acid, activity coefficient  
 of, 198, 228  
 dielectric increment of, 146, 290, 545  
 dipole moment of, 290  
 $pK'$  values of, from cells with liquid  
 junction, 99, 121  
 $K_z$  values of, 99  
 relaxation time of, 566
- $\alpha$ -Aminobutyryl- $\alpha$ -aminobutyric acid,  
 $pK'$  values of, in water, from cells  
 with liquid junction, 84
- $\alpha$ -Amino-*n*-caproic acid\* (norleucine)  
 $\epsilon$ -Aminocaproic acid\*  
 $\alpha$ -Aminocaproic hydantoic acid, appar-  
 ent molal volume of, 372  
 solubility of, in electrolytes in ethanol-  
 water mixtures, 274  
 in ethanol-water mixtures, 215  
 in water and organic solvents, 201,  
 206, 208, 209
- $\epsilon$ -Aminocaproic hydantoic acid, apparent  
 molal volume of, 372  
 solubility of, in ethanol-water mix-  
 tures, 215  
 in water and in organic solvents, 201,  
 215
- $\alpha$ -Amino-iso-caproic acid, see Leucine\*  
 $\omega$ -Amino-*n*-dodecanoic acid,  $pK'$  values  
 of, in water, from cells with liquid  
 junction, 84, 113
- Amino group, structure of, from Raman  
 spectra, 14-9  
 see also: Ammonium groups.
- $\zeta$ -Aminoheptioic acid, dielectric incre-  
 ment of, 545
- $\alpha$ -Amino-iso-valeric acid, apparent molal  
 volume and electrostriction of, 159
- $\alpha$ -Aminophenylacetic acid, formalde-  
 hyde titration constant of, 137
- $\alpha$ -Aminotricarballylic acid,  $pK'$  values  
 of, in water, from cells with liquid  
 junction, 85
- $\alpha$ -Amino-*n*-valeric acid\*  
 $\beta$ -Aminovaleric acid, activity coefficient  
 of, 198, 228
- $\gamma$ -Aminovaleric acid, activity coefficient  
 of, 198, 228  
 in ethanol-water mixtures, 111  
 apparent molal volume and electro-  
 striction of, 159  
 dielectric increment of, 146  
 $pK'$  values of, in water, from cells with  
 liquid junction, 84
- $\delta$ -Amino-*n*-valeric acid, activity coeffi-  
 cients of, in ethanol-water mixtures,  
 111  
 dielectric increment of, 146, 290  
 dipole moment of, 290  
 $K_z$  values of, 99  
 $pK'$  values of, in water, from cells with  
 liquid junction, 84, 99, 121
- Ammonia in proteins, content of, 358,  
 368, 499, 501  
 significance of, for protein titration  
 curves, 444, 499-501
- Ammonium groups, heats of ionization  
 of, 82, 89, 445  
 $pK'$  values of, 80, 84-6, 99, 107, 109,  
 122, 445  
 effect of charged substituents on,  
 117-20  
 effect of dipolar substituents on,  
 120-4  
 effect of resonance on, 124-8  
 structure of, from Raman spectra, 14-9  
 from x-ray diffraction, 318-20
- Ammonium sulfate, activity coefficients  
 of cystine in solutions of, 243-6  
 effect of, on solubility and activity  
 coefficients of proteins, 572, 577,  
 582, 587, 602, 604, 607-8
- Anomalous viscosity, see Non-New-  
 tonian flow
- Anserine,  $pK'$  values of, in water, from  
 cells with liquid junction, 85
- Antibodies, sedimentation constants and  
 molecular weights of, 433
- Antipneumococcus serum globulin,† mo-  
 lecular length of, from frictional  
 ratio and from double refraction of  
 flow, 541
- Apparent molal compressibility of amino  
 acid solutions, 173
- Apparent molal heat capacity of amino  
 acid solutions, 165-73
- Apparent molal properties, of amino  
 acids and related compounds in solu-  
 tion, 155-73  
 of electrolyte solutions, variation with  
 ionic strength, 61

\* See Tabular Index of Amino Acids and Peptides, pp. 680-683.

† See Tabular Index of Proteins, p. 685.



- Apparent molal volumes, as a function of concentration, 61, 62, 161-5  
 definition of, 156-7  
 of amino acids and peptides, 155-65  
 of hydantoic acids, 372  
 of non-electrolytes, 163-5  
 of proteins in solution, 370, 374-7, 428-31
- Apparent specific volume, of amino acid residues, 372  
 of amino acids in solid state and in solution, 376  
 of proteins in solution, 370, 374-7, 428-31  
 of proteins in the solid state, 328-30, 377-81  
*see also*: Partial specific volume
- Arginine\*  
 Ascorbic acid,  $pK'$  values of, 135  
 Asparagine\*  
 Aspartic acid\*  
 Aspartic acid hydantoin, solubility of, in water and in ethanol, 201  
 Aspartylaspartic acid,  $pK$  values of, in water, from cells with liquid junction, 85, 88  
 Aspartylglycine,  $pK'$  values of, in water, from cells with liquid junction, 85  
 $\alpha$ -Aspartylhistidine,  $pK'$  values of, in water, from cells with liquid junction, 86  
 $\beta$ -Aspartylhistidine,  $pK'$  values of, in water, from cells with liquid junction, 86  
 Aspartyltyrosine, heats of ionization of, 89  
 $pK'$  values of, in water, from cells with liquid junction, 85
- Association of ions due to electrostatic forces, 64
- ### B
- Bacillus Phlei protein, molecular weight, sedimentation and diffusion constants and specific volume of, 428  
 Barium bromate, solubility of glycine and alanine in, 253-6  
 Barium iodate, solubility of glycine and alanine in, 253-6  
 Basic amino acids in proteins, 352-6  
 relation of, to maximum acid combining capacity, 354, 446, 478-505  
 Bence-Jones protein  $\alpha$ †  
 Bence-Jones protein  $\beta$ †
- Benzamide, solubility of, in ethanol-water mixtures, 215  
 Bergmann-Niemann hypothesis of protein structure, 315-7  
 Betaine\* (N-trimethylglycine)  
 Betaines, apparent molal volumes of, 159  
 dielectric increments of, 147  
 Raman spectra of, 15  
 transformation of, into isomeric esters, 75
- Boiling point of solvent, effect of solutes on, 32
- Bromelin, action of, on synthetic peptides, 310  
 specificity of, 309-10
- Bromo-acids, dissociation constants ( $pK$  values) of, 121
- Brownian movement, rotary, relation of, to rotary diffusion constants and relaxation times, 396-7, 506-12  
 translational, relation of, to translational diffusion constant, 396-7, 402-6  
*see also*: Diffusion constant, translational; Rotary diffusion constant; Sedimentation; Viscosity; Double refraction of flow; Dielectric dispersion
- Bushy-stunt virus†
- Butanol, extraction of amino acids from protein hydrolysates by, 361  
 solubility of amino acids, peptides and related compounds in, 201, 202, 208, 210, 212
- Butyl alcohol, *see* Butanol
- n*-Butyric acid, apparent molal heat capacities of, 169  
 heats of ionization of, 82, 117
- ### C
- Calcium chloride, activity coefficients of amino acids in solutions of, 240, 243-5  
 denaturing action of, on myosin, 536  
 effect of, on activity coefficients of proteins, 620-1
- Calcium iodate, solubility of glycine and alanine in, 253-6
- Canavalin, diffusion constant, molecular weight, sedimentation constant, and specific volume of, 428
- Caproamide, solubility of, in water and organic solvents, 206
- Carboxyhemoglobin, *see* Hemoglobin†

\* See Tabular Index of Amino Acids and Peptides, pp. 680-683.

† See Tabular Index of Proteins, pp. 684-686.

- Carboxyl group, heat of ionization of, 82, 89, 445  
 ionization of, studied by Raman spectra, 12-5  
 $pK$  values of, 82, 84-6, 99, 107, 109, 117, 121, 445  
 effect of substituents on, 117-24  
 structure of, in amino acids, from x-ray studies, 318-20
- Carnosine,  $pK'$  values of, in water, from cells with liquid junction, 85
- Casein,† molecular weight of, from osmotic pressure, 391  
 solubility of, 586-94  
*see also*: Sodium caseinate
- Catalase†
- Cathepsin, action of, on synthetic peptides, 310
- Centrifugal fields, effect of, on chemical potentials, 48; *see also* Chapter 19.
- $CH_2$  group, *see* Methylene group
- Chemical equilibria, effect of electrostatic interactions on, 67-72  
*see also*: Dissociation constants; Activity coefficients
- Chemical kinetics, effect of electrostatic forces on, 72-4
- Chemical potentials, in electrolyte solutions, 57-60  
*see also*: Activity; Activity coefficients; Free energy
- Chlorocruorin (*Spirographis*), sedimentation constant of, 430
- Chlorophenols,  $pK'$  values of, 130
- Chymotrypsin, crystalline, action of, on synthetic peptides, 310  
 application of phase rule to solubility of, 581-3  
 density of, 328  
 dimensions of unit cell from x-ray data, 328  
 molecular weight of, from osmotic pressure, 390  
 from x-ray diffraction, 328  
 specificity of, 309-10  
 x-ray data on, 328
- $\gamma$ -Chymotrypsin, dimensions of unit cell from x-ray data, 328
- Chymotrypsinogen, application of phase rule to solubility of, 580-1  
 molecular weight of, from osmotic pressure, 390  
 sulfur-containing amino acids of, 348
- Clupein, composition and structure of, 312
- Coconut globulin†
- Collagen, structure of, deduced from x-ray studies, 326-7
- Compressibility of amino acid solutions, 173
- Concanavalin A†
- Concanavalin B†
- CONH— group, influence of, on acidity of a neighboring group, 134, 135  
 on relative solubility in water and organic solvents, 206, 211, 212
- Conjugated proteins, 307-8
- Cosubstrates in proteolysis, 314
- Couette viscometer, 513-4
- Covolume (Traube), definition and magnitude of, for organic substances in aqueous solution, 157, 370-4
- Creatine, acidity of, 133  
 dielectric increment of, 146
- Critical frequency,  $\nu_c$ , definition of, and relation to relaxation time, 555  
*see also*: Relaxation time
- Crotoxin†
- Cyclol hypothesis of protein structure, 334-6
- Cysteine, in proteins, 346-52  
 $pK'$  values of, in water, from cells with liquid junction, 85
- Cysteinylcysteine,  $pK'$  values of, in water, from cells with liquid junction, 85
- Cystine,\* in proteins, 346-52  
 salting-out constants of, 604
- Cystinyldiglycine, dielectric increment of, 146  
 $pK'$  values of, in water, from cells with liquid junction, 86
- Cystinyldiglycine, dielectric increment of, 146  
 $pK'$  values of, in water, from cells with liquid junction, 86
- Cytochrome-c,† titration curve, structure and heme-linked acid groups of, 486-9

## D

- Deamination of proteins, effect of, on gelatin, 492-5  
 on protein titration curves, 448, 492-5
- Debye-Hückel theory, of electrolyte solutions, application of, to acid-base equilibria, in amino acids, 105-11; in proteins, 468-77  
 derivation of fundamental equations, 52-9

\* *See* Tabular Index of Amino Acids and Peptides, pp. 680-683.

† *See* Tabular Index of Proteins, pp. 684-685.

- see also*: Electrostatic forces; Activity coefficients; Ionic strength; Dielectric constant; Solubility of electrophoretic migration, 642-3
- Denaturation of proteins, methods of avoiding, during extraction and purification, 570-3
- observed by alterations, of diffusion constant, 417-8
- of double refraction of flow, 536-41
- of osmotic pressure, 389-92
- of sedimentation constant, 435-43
- of thickness of surface layers, 436
- of titratable sulphhydryl groups, 435, 437, 537
- of viscosity, 520-22, 526-7
- of x-ray diffraction patterns, 333
- theory of, 437
- Density and specific volume of proteins, 328-32, 370-81, 428-31
- Dextrose, apparent molal heat capacity of, 169
- $\epsilon'$ -Diamino-di ( $\alpha$ -thio-*n*-caproic) acid, apparent molal volume and electrostriction of, 159
- $\alpha$   $\beta$ -Diaminopropionic acid,  $pK'$  values of, in water, from cells with liquid junction, 85
- Dibromo-*L*-tyrosine, apparent heats of ionization of, 89
- $pK'$  values of, in water; from cells with liquid junction, 85
- presence of, in proteins, 343
- solubility of, in ethanol-water mixtures, 216
- in water, 200
- at different temperatures, 190
- Dicarboxylic acids, in proteins, 358, 368, 499, 501
- see also*: Aspartic acid; Glutamic acid; Hydroxyglutamic acid
- Dichloro-*L*-tyrosine, apparent heats of ionization of, 89
- $pK'$  values of, in water, from cells with liquid junction, 85
- solubility of, in ethanol-water mixtures, 216
- in water, 200
- at different temperatures, 190
- Dichlorophenols,  $pK'$  values of, 130
- Dielectric constant, definition of, 140
- methods of measurement of, 140, 547-9
- relation of, to dipole moments, in gases and nonpolar solvents, 142-4
- in polar solutions, 149-52, 294-6, 546-7
- significance of, for activity coefficients, in electrolyte solutions, 52-74, 236
- in solutions of dipolar ions, 217-87
- in extraction and purification of proteins, 571
- variation of, with frequency, *see* Dielectric dispersion
- see also*: Dielectric increment
- Dielectric dispersion, 543-58
- in amino acid and peptide solutions, 565-7
- in various protein solutions, 560-5
- inferences from, concerning protein structure, 567-8
- relation of, to dispersion of conductance, 558-60
- to molecular size and shape, 555-8
- theory of, 543-6, 555-8
- see also*: Dielectric constant; Dielectric increment; Dipole moment; Relaxation time
- Dielectric increment, definition of, 145
- effect on, of rotation around valence bonds, 152-4
- high-frequency increment, 553-4
- low-frequency increment, 550-3
- relation of, to dipole moment, 149-52, 290, 294-6, 546-7
- table of values of, for amino acids at several temperatures, 148
- for amino acids, peptides, and betaines, 146
- for  $\alpha$ -aminobutyric acid in various solvents, 148
- for proteins in solution, 545, 554, 557, 559
- for uncharged molecules, 144
- total increment, 554-5
- variation with frequency, *see* Dielectric dispersion; Relaxation time
- Diethyl aspartate,  $pK$  values of, in water, 99
- Diethyl glutamate,  $pK$  values of, in water, 99
- Diffusion, rotary, *see* Rotary diffusion constant
- Diffusion, translational, general principles and laws of, 397-407
- modification of, by a centrifugal field, 419
- see also*: Diffusion constant
- Diffusion coefficient, *see* Diffusion constant
- Diffusion constant, rotary, *see* Rotary diffusion constant
- Diffusion constant, translational, definition of, 397-8

- Diffusion constant, translational, experimental methods for determining, 406-10  
 relation of, to Brownian movement, 402-6  
   to molecular size and shape, 402-6, 419-25  
   values of, for amino acids, 410-12  
   for proteins, 411-8, 428-31
- Diglycine hydantoic acid, apparent molal volume of, 372  
 solubility of, in electrolytes in ethanol-water mixtures, 274  
 in water and in organic solvents, 201, 206
- Diglycylcystine, dielectric increment of, 146  
 $pK'$  values of, in water, from cells with liquid junction, 86
- 3,4-Dihydroxyphenylalanine,  $pK'$  values of, in water, from cells with liquid junction, 85
- Diiodotyrosine, apparent heats of ionization of, 89  
 heats of solution of, 192, 194  
 $pK'$  values of, in water, from cells with liquid junction, 85  
 presence of, in proteins, 342  
 solubility of, in ethanol-water mixtures, 216  
 in water, 200  
   at different temperatures, 190
- Diketopiperazine, effect of, on solubility of  $\text{TiCl}_4$  and  $\text{TiIO}_3$ , 254  
 structure of molecule and crystal of, 318-9
- Diketopiperazine rings, lack of evidence for their presence in proteins, 310
- N*-Dimethyl-*o*-aminobenzoic acid,  $K_z$  values of, 99  
 $pK'$  values of, in water, from cells with liquid junction, 99
- N*-Dimethylglycine, activity coefficient of, 254  
 $pK'$  values of, in water, from cells with liquid junction, 85
- N*-Dimethylphenylglycine, apparent molal volume and electrostriction of, 159
- Diphtheria antitoxic pseudoglobulin†
- Diphtheria antitoxin, solubility and purification of, 585  
 molecular weight, sedimentation constant, diffusion constant, partial specific volume of, 428
- Diphtheria toxin, molecular weight, sedimentation constant, diffusion constant, partial specific volume of, 428
- Dipolar ionic structure, effect of, on apparent molal compressibilities, 173  
 on apparent molal heat capacities, 165-73  
 on apparent molal volumes, 157-65  
 on crystal density, 186  
 on crystal structure, 184-7  
 on dielectric increments and dipole moments, 145-52, 550-68  
 on infrared and ultraviolet spectra, 19  
 on melting points, 184-5  
 on  $pK'$ , as a function of dielectric constant of solvent, 105-11  
   as a function of ionic strength, 111-5  
 on Raman spectra 13-9  
 on solubility and activity coefficients 187 ff.  
   in electrolyte solutions 236-75  
   277-87 609-17  
   in solutions of other dipolar ions, 217-35, 617-22  
 on solubility ratios and activity coefficients in water and organic solvents, 206, 208, 209, 212-14  
 on surface tension of amino acid solutions, 174-6
- Dipolar ions, equilibrium of, with uncharged molecules, 75, 96-115, 127-9  
 effect of change of solvent on, 105-11  
 effect of temperature on, 104  
 history of discovery of, 1-8
- Dipole-dipole interactions, Fuoss's theory of, 170, 171, 232, 234  
 in amino acid and peptide solutions, 217-35  
 in protein solutions, 617-22  
 Kirkwood's theory of, 234  
 Scatchard-Kirkwood "dumb-bell" model for, 108-10, 231-5
- Dipole moment, electric, definition of, 2, 142-4  
 influence of, on activity coefficients, in interaction of dipolar ions, 226 ff., 617-9  
 in interaction of ions and dipolar ions, 272-5, 277-87  
 of a substituent, influence of, on ionization of a neighboring group, 117-24, 287-90

† See Tabular Index of Proteins, pp. 684-685.

- relation of, to dielectric increment, 149-52, 290, 294-6  
 table of values of, for certain polar molecules, 141, 144  
 tables of values of, for amino acids and peptides, 146, 148, 290  
 for proteins, 554  
*see also*: Dielectric constant; Dielectric increment; Dielectric dispersion; Electric moments
- Dispersion of conductance, relation of, to dielectric dispersion and relaxation times, 558-60
- Dispersion of dielectric constant, *see* Dielectric dispersion
- Dissociation and association of proteins, observed by changes in osmotic pressure, 391  
 observed by changes in sedimentation constant, 437-43  
*see also*: Denaturation of proteins
- Dissociation constants (acidic), effect of peptide linkage on, 134-5  
 effects of electrostatic forces on, 117-24, 287-94, 468-77  
 effects of resonance on, 124-30  
 of acidic and basic groups in proteins, 445  
 of amino acids, 75-139  
   in ethanol-water mixtures, 105-11  
   in formaldehyde solutions, 136-9  
   in water, from cells with liquid junction, 83-90  
     from cells without liquid junction, 75-83  
 of carboxyhemoglobin estimated from solubility measurements, 595-8  
 of certain ring structures, 135  
 of guanidine derivatives, 132-3  
 of hydantoin, 135  
 of imidazole derivatives, 131-2  
 of ionizing groups in proteins, 445, 465-505  
 of organic derivatives of phosphoric acid, 133-4  
 of phenol derivatives, 84-5, 130  
 of pyrazole derivatives, 131-2  
 of substituted groups, 85, 131
- Donnan equilibrium, 46  
 influence of, on osmotic pressure of proteins, 395
- Double refraction of ellipsoidal molecules in parallel orientation, 531-2
- Double refraction of flow, 527-42  
 first discovery of, 527-8
- in myosin, 528, 536-7, 540  
 in polydisperse systems, 539-41  
 in tobacco mosaic virus, 537-9, 540  
 in  $V_2O_5$  sols, 528  
 molecular lengths determined from, 536-42  
   comparison of, with lengths calculated from sedimentation and diffusion, 541-2  
   theory of, 532-6
- "Dumb-bell" model of a dipolar ion, use of, in calculation of thermodynamic properties, 108-10, 231-5

## E

- Edestin,† asymmetry and hydration of molecule of, 562  
 dipole moment of, 554  
 solubility of, 599-600
- Egg albumin,† asymmetry and hydration of molecule of, 562  
 conductance increment of, 559  
 dipole moment of, 554  
 heat of solution of, 587  
 influence of ionic strength on titration curve of, 469  
 salting-out constant of, 604  
 solubility of, 577, 587, 606
- Einstein's equation for viscosity of solutions of spherical molecules, 515-6
- Einstein's theory of diffusion and Brownian movement, 402-6, 510
- Electric double layer, 52 ff., 625 ff.
- Electric moments, influence of, on activity coefficient of a spherical dipolar ion with any given charge distribution, 609-17; *see also* 277-87  
*see also*: Dipole moment, electric
- Electrode polarization in dielectric constant measurements, 549
- Electromotive force measurements, on cells with liquid junction, in amino acid solutions, 83-90; also 96-135  
 in protein solutions, 453-7; also 469-505  
 on cells without liquid junction, determination of dissociation constants of amino acids and other acids by, 79-84  
 determination of interaction between amino acids and salts by, 251-2, 256-65  
 between proteins and salts by, 619-22

† *See* Tabular Index of Proteins, pp. 684-685.

- Electromotive force measurements, on cells without liquid junction, determination of maximum acid combining capacity of proteins by, 457-60  
thermodynamic principles of, 37-45  
*see also*: Dissociation constants; Activity coefficients
- "Electrophoretic" friction force, 625 ff.
- Electrophoretic migration, effect of ionic strength on, 640-3  
of pressure differences in liquid on, 629, 632  
flow parallel to plane surfaces, 626-31  
of spherical particles, 631-43  
theory of, 623-43
- Electrostatic forces, influence of, on acid-base equilibria, 117-24, 287-94, 468-77  
on apparent molal properties of ions and dipolar ions, 155-61, 165-73  
on properties of solutions of dipolar ions, 277-87  
of solutions of ions, 52-74, 236  
on solubility of proteins, 609-17  
*see also*: Activity coefficients; Electric moments
- Electrostriction, as a factor in calculation of volumes of amino acid residues, 373  
of proteins in solution, 378, 517  
of solvent by charged groups, in amino acid solutions, deduced from apparent molal volumes, 158-61  
in ionic solutions, 155, 156, 160, 161  
relation of, to apparent molal heat capacities, 165-73
- Electroviscous effects, 523
- Entropy, 21 ff.  
decrease of, on formation of a dipolar ion in solution, 187  
of solution, in aqueous solutions, 182-3  
in regular solutions, 180
- Equilibria between ions and dipolar ions, in amino acid solutions, 75-115  
in protein solutions, 460-77
- Equilibrium, chemical, general thermodynamic conditions for, 29, 36
- Erythrocrucorin† (Arca, Arenicola, Chironomus, Daphnia, Lampetra, Lumbricus, Planorbis)
- Ethanol, solubility of amino acids in, 198-216
- Ethanol-water mixtures, dielectric increment of  $\alpha$ -aminobutyric acid in, 148  
dissociation constants of amino acids in, 106-11  
effect of, on titration curves of proteins, 447, 489-95  
solubility of amino acids in, 203, 204  
in presence of electrolytes, 266-75  
Ethyl alcohol, *see* Ethanol
- Ethylammonium ion, *pK* values of, in ethanol-water, 109  
in water, 109
- Ethyl aspartate, *pK* values of, in water, 99
- Ethyl glutamate, *pK* values of, in water, 99
- $\alpha$ -Ethyl hydrogen glutamate, *pK* values of, in water, 99
- $\gamma$ -Ethyl hydrogen glutamate, *pK* values of, in water, from cells with liquid junction, 99
- Euglobulin, molecular weight of, from osmotic pressure, 390
- Excelsin, † x-ray data on, 328
- Extraction of proteins, significant factors in, 569-73

## F

- Ferrihemoglobin* (methemoglobin), heme-linked acid groups of, 482-6
- Fibrin†
- Fibrin (cattle)†
- Fibrinogen, rotary diffusion constant and approximate length of, 540  
salting-out constant of, 604  
solubility of, 602
- Fibrous proteins, structure of, deduced from x-ray diffraction studies, 322-7
- Fick's laws of diffusion, 397-8
- Formaldehyde; effect of, on gelatin titration curve, 492-5  
on protein titration curves, 447, 492-5  
on titration constants of amino acids, 136  
titration of amino acids with, principles underlying, 136-9
- Formamide, solubility of amino acids, peptides and related compounds in, 201, 202, 208, 210, 212
- Formic acid, apparent molal heat capacity of, 169  
heats of ionization of, 82, 117
- Formyl- $\alpha$ -aminobutyric acid, solubility of, in ethanol-water mixtures, 215  
in water and organic solvents, 201, 202, 206, 208, 209

† *See* Tabular Index of Proteins, pp. 684-685.

- Formylglycine, melting point of, 185  
 $pK'$  value of, 135  
 solubility of, in ethanol-water mixtures, 215  
   in water and organic solvents, 201, 202, 206, 208, 209
- Formylleucine, melting point of, 185  
 solubility of, in ethanol-water mixtures, 215  
   in water and organic solvents, 201, 206
- Free energy, of solution, influence of  $\text{CH}_2$  groups on, 212  
*see also*: Activity coefficients
- Freezing point of solvent, effect of solutes on, 32  
 study of interaction of amino acids and salts by, 251, 256-64
- Frictional ratio ( $f/f_0$ ), definition of, 404-6  
 relation of, to hydration, 424-5, 434-7  
   to molecular shape, 405-6, 424-5, 434-7  
 values of, for protein molecules, 428-31, 520, 541
- G
- Gelatin† deamination of, effect of, on titration curve, 494-5  
 double refraction of, as a function of refractive index, 532  
 effect of formaldehyde on titration curve of, 492-5  
 interaction of, with salts, 620, 621
- Gibbs-Duhem equation, 21
- Gliadin,† dipole moment of, 554
- Globin: sheep, ox, dog: molecular weight of, from osmotic pressure, 391  
 "Globular" proteins, crystal structure and size of unit cell of, 327-32  
 molecular weights and shape of, 428-43
- Glutamic acid\*  
 Glutamine\*  
 Glutaminylglutamic acid,  $pK'$  values of, in water, from cells with liquid junction, 85  
 Glutaminylglycine,  $pK'$  values of, in water, from cells with liquid junction, 85  
 Glutathione,  $pK'$  values of, in water, from cells with liquid junction, 85  
 splitting of, by carboxypeptidase, 312
- Glycine,\* comparison of, with its isomers glycylamide and nitroethane, 188  
 interatomic distances in, 320  
 solubility of proteins in, 587, 617-9
- Glycine amide,  $pK'$  values of, in water, from cells with liquid junction, 84
- Glycine residue, solubility ratios of, in organic solvents, 212
- Glycolamide, apparent molal heat capacity of, 170, 171  
 apparent molal volume of, 158, 163  
 comparison of, with its isomers glycine and nitroethane, 188  
 density of solid, 186  
 melting point and solubility of, 185, 188  
 solubility of, in water and organic solvents, 201, 202, 206
- Glycolylglycineamide, solubility of, in water and organic solvents, 201, 206
- Glycylalanine, activity coefficient of, 198, 228  
 apparent molal volume and electrostriction of, 159  
 dielectric increment of, 146  
 heats of ionization of, 89  
 $pK'$  values of, in water, from cells with liquid junction, 84  
 relaxation time of, 566
- Glycylalanylalanylglycine,  $pK'$  values of, in water, from cells with liquid junction, 84
- Glycyl- $\alpha$ -amino-tricarballic acid,  $pK'$  values of, in water, from cells with liquid junction, 85
- Glycylaspartic acid,  $pK'$  values of, in water, from cells with liquid junction, 85
- Glycylglycine\*  
 Glycylglycine ester,  $pK$  values of, in water, from cells with liquid junction, 99, 113
- Glycylleucine, apparent molal volume and electrostriction of, 159  
 dielectric increment of, 146  
 $pK'$  values of, from cells with liquid junction, 84
- Glycylphenylalanine, apparent molal volume and electrostriction of, 159  
 dielectric increment of, 146
- Glycylproline,  $pK'$  values of, in water, from cells with liquid junction, 84
- Glycyl-sarcosine,  $pK'$  values of, in water, from cells with liquid junction, 85
- Glycyltyrosine,  $pK'$  values of, in water, from cells with liquid junction, 85

\* *See Tabular Index of Amino Acids and Peptides*, pp. 680-683.  
 † *See Tabular Index of Proteins*, pp. 684-685.

Glycylvaline,  $pK'$  values of, in water, from cells with liquid junction, 84  
 Gravitational fields, effect of, on chemical potentials, 48  
 Guanidine hydrochloride, effect of, on diffusion constants and viscosity of proteins, 437, 527  
   on titratable sulfhydryl groups in proteins, 437  
   on viscosity and double refraction of myosin, 527, 537  
 Guanidinium ion, acidity of, and its derivatives, 84-6, 132-3, 445  
   heats of ionization of, in guanidine derivatives, 89, 445  
   structure of, 132  
 Güntelberg-Müller charging process, 66

## H

Heat capacity, of electrolyte solutions, 165-7  
   of amino acid solutions, 165-73  
 Heats of ionization, of acid and basic groups, in cytochrome-c, 488  
   in egg albumin, 498-9  
   in hemoglobin, 481-2  
   in insulin, 502  
   in  $\beta$ -lactoglobulin, 500  
   in proteins, 445, 447  
   in serum albumin, 503-5  
   in wool protein, 496  
   in zein, 491  
   of amino acids, 81-3, 88-90  
   of ammonium groups, 82  
   of carboxyl groups, 82  
   of chloroacetic acid, 117  
   of fatty acids, 117  
   of phenolic groups, 82  
   relation of, to solubility, 590-3  
 Heats of solution, of alcohols, ethers and ketones, in water, 182-3  
   of amino acids and peptides in water, 187-95  
   of proteins, in water, 586-8  
   of racemic and optically active amino acids, 194  
   of the components of regular solutions in one another, 180  
   table of values of, for amino acids, 192  
   for proteins, 587  
 Heme-linked acid groups, of cytochrome-c, 486-9  
   of *ferro*, *ferri*, and oxyhemoglobin, 482-6

Hemocyanin,† dissociation and association of, studied by ultracentrifugal measurements, 437-43  
 Hemocyanin, *Busycon*†  
 Hemocyanin, *Calcaris*†  
 Hemocyanin, *Eledone*†  
 Hemocyanin, *Helix*,† approximate length of, 540-1  
 Hemocyanin, *Helix nemoralis*†  
 Hemocyanin, *Helix pomatia*,† pH stability diagram of, 440  
 Hemocyanin, *Homarus*†  
 Hemocyanin, *Limulus polyphemus*,† pH stability diagram of, 438  
 Hemocyanin, *Nephrops*†  
 Hemocyanin, *Palinurus vulgaris*,† pH stability diagram of, 438  
 Hemocyanin, *Pandalus*†  
 Hemocyanin, *Rossia*†  
 Hemoglobin, cattle, acid and base binding capacity of, 478  
   density of crystal of, 378  
 Hemoglobin, horse,† apparent heat of dissociation of groups in, 481  
   asymmetry and hydration of molecule of, 562  
   conductance increment of, 559  
   dipole moment of, 554  
   distribution among forms of different net charge in isoelectric, 467  
   heme-linked acid groups of, 482-6  
   interaction of, with glycine, 233, 617-8  
   with salts, determined from EMF measurements, 620  
   salting-out constant of, 604  
   solubility and heat of solution of, 587  
   solubility of, 596, 597, 602-13, 617-8  
 Hemoglobin, man,† salting-out constant of, 604  
 CO-Hemoglobin, *Lampetra*, diffusion constant of, 416  
 Hemoglobin, pig, dielectric increment and dipole moment of, 554  
   relaxation time and critical frequency of, 557  
 Hemoglobin, sheep, molecular weight of, from osmotic pressure, 392  
 Henry's theory, of electrophoretic migration, 636-43  
 Heptaglycine, dielectric increment of, 146, 545  
 Heptanol, solubility of amino acids, peptides and related compounds in, 201, 202, 208, 210, 212

† See Tabular Index of Proteins, pp. 684-686.



- Hexaglycine, dielectric increment of, 146  
 $pK'$  values of, in water, from cells with liquid junction, 84
- Hexose-phosphoric acids,  $pK'$  values of, 134
- Histidine\*
- Histidylglycine, apparent heats of ionization of, 89  
 $pK'$  values of, in water, from cells with liquid junction, 85
- Histidylhistidine,  $pK'$  values of, in water, from cells with liquid junction, 85, 87
- Hofmeister series, 64, 621
- Hordein†
- Horse antibody globulin, approximate length of, 540
- Hückel's theory of electrophoretic migration, 636-43
- Human tuberculosis bacillus protein†
- Hydantoic acid, apparent molal volume of, 372  
 melting point of, 185  
 $pK'$  value of, 135  
 solubility of, in ethanol-water mixtures, 215  
 in ethanol-water mixtures and in electrolytes, 274  
 in water and organic solvents, 201, 202, 206, 208, 209
- Hydantoin,  $pK'$  value of, 135  
 solubility of, in ethanol-water mixtures, 215  
 in water and organic solvents, 201, 202
- Hydantoin of alanine, melting point of, 185
- Hydantoin of  $\alpha$ -methyl- $\gamma$ -butyric acid, melting point of, 185  
 solubility of, in water and ethanol-water, 201, 215
- Hydantoin of aspartic acid, melting point of, 185
- Hydantoin of glycine, melting point of, 185
- Hydantoin of leucine, melting point of, 185  
 solubility of, in water and in ethanol, 201
- Hydration, of protein crystals, 328-31, 379-81  
 of proteins in solution, 388, 394-5, 405-6, 421-5, 434, 516-7, 519-22, 553, 560-5
- Hydrogen bonds, definition of, 181  
 significance of, for crystal structure, of alanine, 186-7  
 of diketopiperazine, 318-9  
 of glycine, 186-7  
 for protein structure and denaturation, 437 (footnote)  
 for structure of water and ice, 160, 181
- Hydroxyamino acids, as constituents of proteins, 340-2
- Hydroxyamides, see Glycolamide, Lactamide,  $\alpha$ -Hydroxycaproamide, Glycylglycine-amide
- $\alpha$ -Hydroxyasparagine,  $pK'$  values of, in water, from cells with liquid junction, 84
- $\beta$ -Hydroxyasparagine,  $pK'$  values of, in water, from cells with liquid junction, 84
- $\alpha$ -Hydroxycaproamide, solubility of, in water and organic solvents, 201, 206
- Hydroxyglutamic acid, content of, in proteins, 358, 368  
 isolation and possible structure of, 342
- $\beta$ -Hydroxyglutamic acid,  $pK'$  values of, in water, from cells with liquid junction, 85
- Hydroxyl group (aliphatic), influence of, on acidity of a neighboring group, 121, 122  
 on activity coefficients of amino acids in aqueous solutions, 221, 222  
 on relative solubility in water and organic solvents, 206, 211, 212
- Hydroxyl group (phenolic), acidic strength of, 130  
 heat of ionization of, 82, 89, 445
- Hydroxylysine, content of, in proteins, 342, 369
- Hydroxyproline\* (oxyproline)
- Hydroxyvaline (oxyvaline), heats of ionization of, 89  
 $pK'$  values of, in water, from cells with liquid junction, 84

## I

- Imidazole derivatives,  $pK'$  values of, 131; see also 84-6
- Inorganic salts, melting points of, 185
- Insulin,† dipole moment of, 554  
 solubility and heat of solution of, 587  
 solubility of, 575, 601

\* See Tabular Index of Amino Acids and Peptides, pp. 680-683.

† See Tabular Index of Proteins, pp. 6-4-656.

- Insulin, structure of (Patterson projections), 336  
 x-ray data on, 328
- Interatomic distances, in amino acids and peptides, 318-21
- Iodination of amino acids and proteins, 343, 448, 491-2
- Iodoproteins, 342-4
- Iodozein, acid and base binding of, 491
- Ional concentration, definition of, 54.  
*See also:* Ionic strength.
- Ionic strength, definition of, 54  
 effect of, on acid-base equilibria, in amino acid and peptide solutions, 111-5, 287-94  
 in protein solutions, 287-94, 468-77  
 on activity coefficients, of amino acids and peptides, 236-87  
 of dipolar ions, 111-5, 236-87  
 of ions, 57-60, 66-72, 111-5, 236-7, 471-7  
 of proteins, 468-77, 602-22  
*see also:* Salting-in coefficient; Salting-out
- on apparent and partial molal properties of electrolytes, 60-2, 161-3  
 on electrophoretic migration, 639-43  
*see also:* Debye-Hückel theory; Ional concentration
- Isoasparagine ( $\alpha$ -amide),  $pK'$  values of, in water, from cells with liquid junction, 84
- Isoelectric point, of amino acids and peptides, relation of, to  $pK$  values, 90-3  
 table of values of, 84-6  
 of proteins, distinction of, from isoelectric point, 446  
 relation of, to solubility, 573  
 relation of, to dissociation constants of ampholytes, 90-3
- Isoglutamine ( $\alpha$ -amide),  $pK'$  values of, in water, from cells with liquid junction, 84
- Isoionic point, of proteins in solution, 446, 468-71
- Isoleucine\*
- Ioserine,  $pK'$  values of, in water, from cells with liquid junction, 84
- K**
- Keratin (wool),<sup>†</sup>  $\alpha$ ,  $\beta$  and supercontracted forms of, 325
- structure of fibers of, deduced from x-ray studies, 323-6  
*see also:* Wool protein
- $K_z$ , equilibrium constant for interconversion of dipolar ions and uncharged molecules,  
 definition and evaluation of, 96 ff.  
 effect of change in medium on, 105-11  
 effect of temperature on, 104-5  
 relation to chemical structure, in alicyclic amino acids, 128  
 in aliphatic amino acids and peptides, 100-4  
 in aminobenzoic acids, 98, 127-9  
 tables of values of, 99, 107, 128  
*see also:* Dissociation constants; Acid base equilibria; Dipolar ionic structure
- L**
- Lactalbumin<sup>†</sup>
- Lactamide, apparent molal heat capacity of, 171, 172  
 apparent molal volume of, 158, 163  
 melting point of, 185  
 solubility of, in water and organic solvents, 201, 202, 206
- $\beta$ -Lactoglobulin<sup>†</sup> (Lactoglobulin), asymmetry and hydration of, 562  
 comparison of analytical and electromotive force studies on, 500-1  
 conductance increment of, 559  
 dicarboxylic acids of, 500-1  
 dipole moment of, 554  
 effect of ionic strength on titration curve of, 470  
 solubility of, in glycine solutions, 587, 619  
 in salt solutions, 587, 610, 614  
 in water, and heat of solution, 587, 614, 619  
 x-ray data on crystals of, 328
- Lead iodate, solubility of glycine and alanine in, 253-6
- Leucine\* ( $\alpha$ -amino-isocaproic acid), apparent molal volume and electrostriction of, 159  
 salting-out constants of, 604
- Leucylalanine, relaxation time of, 566
- Leucylglycine, apparent molal volume and electrostriction of, 159  
 dielectric increment of, 146  
 relaxation time of, 566

\* See Tabular Index of Amino Acids and Peptides, pp. 680-683.

† See Tabular Index of Proteins, pp. 684-686.

Leucylglycylglycine (Leucylidiglycine),  
 apparent molal volume and electro-  
 striction of, 159  
 dielectric increment of, 146  
 diffusion constant of, 412  
 relaxation time of, 566  
 Leucyl octadecylglycylglycine,  $pK'$  values of,  
 in water, from cells with liquid junc-  
 tion, 84  
 Linderström-Lang's equation for equilib-  
 rium between ions and dipolar ions  
 in ampholyte (protein) solutions,  
 70-2, 463  
 Linderström-Lang's theory of effect of  
 ionic strength on protein titration  
 curves, 470-7  
 Liquid junction potentials, 39-45, 83,  
 453  
 stabilization of, by addition of large  
 excess of neutral salt, 457  
 Lithium chloride, activity coefficients  
 of amino acids in solutions of, 242,  
 269-75  
 Lysine\*  
 Lysylglutamic acid\*  
 Lysyl lysine, apparent heats of ioniza-  
 tion, 89  
 $pK'$  values of, in water, from cells  
 with liquid junction, 85

M

Magnesium chloride, effect of, on ac-  
 tivity coefficients of proteins, 620-1  
 Magnesium sulfate, effect of, on activity  
 coefficients of proteins, 572, 581  
 Mass action law, 21  
 Maximum acid and base combining ca-  
 pacity of proteins, calculated from  
 EMF of cells without liquid junc-  
 tion, 351, 457-60  
 methods of determining, 453-60  
 relation of, to structure and compo-  
 sition, 354, 446  
 values of, for specific proteins, see  
 references in *Table of Index of  
 Proteins*  
 Maximum charge, pH of, relation of, to  
 dissociation constants of amphi-  
 lytes, 93-6  
 Melting points, of aliphatic amines and  
 acids, 185  
 of amino acid derivatives, 185  
 of amino acids, 1, 185-185  
 relation of, to dipolar ionic struc-  
 ture, 4, 184

of inorganic salts, 185  
 of normal hydrocarbons, 185  
 of organic compounds, 185  
 Membrane potentials, 47  
 Mesocystine, solubility of, in water, 200  
 at different temperatures, 190  
 Methanol, solubility of amino acids,  
 peptides and related compounds in,  
 201, 202, 208, 210, 212  
 Methemoglobin, horse, density and  
 molecular weight of, from x-ray  
 diffraction, 328  
 horse, x-ray data on, 328  
 sheep, ox, dog; molecular weight of, from  
 osmotic pressure, 390, 391  
 see also: Ferrihemoglobin  
 Methionine\*, in proteins, 346-52  
 Methyl histidine,  $pK'$  values of, in water,  
 from cells with liquid junction, 85  
 Methylene ( $\text{CH}_2$ ) group, influence of, on  
 activity coefficients in aqueous solu-  
 tion, 219, 221, 225  
 on apparent molal volumes, 157-9  
 on apparent molal heat capacities,  
 168-73  
 on interactions between dipolar ions,  
 229  
 on salting-out in amino acid solutions,  
 238-9  
 on solubility ratios and activity coef-  
 ficients, 205-212, 221  
 on surface tension of amino acid solu-  
 tions, 174-6  
 Methylhydantoic acid, apparent molal  
 volume of, 158  
 solubility of, in water and ethanol-  
 water mixtures, 215  
 Mobility, electrophoretic, see Electro-  
 phoretic migration  
 of ions, relation of, to liquid junction  
 potentials, 39-45  
 Molecular weights of proteins, from os-  
 motic pressures, 382-95  
 from sedimentation and diffusion,  
 419-43  
 from sedimentation equilibrium, 420-  
 1, 428-31  
 from size of unit cells in protein crys-  
 tals, 328-31  
 theory of molecular weight classes  
 (Svedberg), 431-33  
 Myogen†  
 Myoglobin, † dipole moment of, 554  
 salting-out constant of, 604  
 solubility of, 602

\* See *Tabular Index of Amino Acids and Peptides*, pp. 680-683.  
 † See *Tabular Index of Proteins*, pp. 694-696.

Myosin,† structure of fibers of, deduced from x-ray studies, 326. See below.  
 Myosin, octopus, rotary diffusion constant and approximate length of, 540  
 Myosin, rabbit,† double refraction of, as a function of refractive index, 532  
   double refraction of flow in, 530  
   molecular weight of, from osmotic pressure, 392  
   rotary diffusion constant and approximate length of, 540  
   sulfur containing amino acids of, 348, 351  
   viscosity of, 525-7  
 Myosin, snail, rotary diffusion constant and approximate length of, 540

## N

Nitroethane, comparison of, with its isomers glycine and glycolamide, 188  
 Non-Newtonian flow, relation of, to molecular asymmetry, 512, 523-7  
   *see also*: Viscosity; Viscosity increment; Rotary diffusion constants  
 Non-polar molecules, 1, 141  
 Norleucine\* ( $\alpha$ -amino-*n*-caproic acid)  
 Normal hydrocarbons, melting points of, 185  
 Nucleohistone (calf thymus)†

## O

Organic solvents, use and dangers of, in protein preparations, 571  
 Ornithine, dielectric increment of, 146  
    $pK'$  values of, from cells without liquid junction, 80  
 Osmotic coefficient, definition of, 26  
   relation of, to activity coefficients, 26  
   in electrolyte solutions, 59-60  
   values of, in serum albumin solutions, 387-8  
 Osmotic pressure, of proteins, 382-95  
   determination of molecular weight from, 382-95  
   in aqueous solutions, 386-90, 392-5  
   in urea solutions, 391  
   influence of activity coefficients of components on, 394  
   thermodynamic principles determining, 45-7  
 Ovalbumin†, *see* Egg albumin†  
 Oxy- $\alpha$ -aminobutyric acid,  $pK'$  values of, in water, from cells with liquid junction, 84  
   apparent heats of ionization of, 89

Oxyglutamic acid, molal volumes and specific volume of residue of, 372  
 Oxyhemoglobin, ox, dog, sheep: molecular weight of, from osmotic pressure, 390, 391  
 Oxyproline\* (Hydroxyproline)  
 Oxytocic pressor hormone, molecular weight, sedimentation and diffusion constant, partial specific volume of, 428  
 Oxyvaline, *see* Hydroxyvaline.

## P

Papain, action of, on synthetic peptides, 310  
   specificity of, 309-10  
 Partial molal quantities, in electrolyte solutions, general equations determining, 61  
   *see also*: Apparent molal volume, heat capacity, and compressibility.  
 Partial specific volumes of proteins in solution, 374-77, 428-31  
   *see also*: Apparent specific volumes; Apparent molal volumes  
 Patterson projections, use of, in study of x-ray data on proteins, 336  
 Pentaglycine, dielectric increment of, 146  
    $pK'$  values of, in water, from cells with liquid junction, 84  
   relaxation time of, 566  
 Pepsin†, action of, on synthetic peptides, 310  
   solubility of, 579  
   specificity of, 309-10  
   x-ray data on crystals of, 328  
   viscosity increment of, 520  
 Peptide linkage, effect of, on acidic dissociation, 134-5  
   evidence for, in protein structure; synthesis of, by proteolytic enzymes, 309-12, 314  
   splitting of, by proteolytic enzymes, 309-15  
 Peptides, activity coefficients of, in electrolyte solutions, 273-5  
   in water and in solutions of other dipolar ions, 217-35  
   apparent molal volumes and electrostriction of, 157-65  
   dissociation constant of, from cells with liquid junction, 83-90  
   distribution of charged forms in, 465

\* See Tabular Index of Amino Acids and Peptides, pp. 680-683.

† See Tabular Index of Proteins, pp. 684-685.

- heats of ionization of, 89  
isoelectric points of, 84-6, 90-3  
partial specific volumes of, 157-65, 376  
solubility of, in water and organic solvents, 198-216  
splitting of, by proteolytic enzymes, 310
- pH, definition of, 43  
*see also*: Dissociation constants (acidic); Electromotive force measurements; Liquid junction potentials; Activity coefficients
- Phase rule, 21  
application of, to solubility of proteins, 576-85
- Phenolic groups, heats of ionization of, 82
- Phenyl group, influence of, on acidity of a neighboring group, 124-9  
on relative solubility in water and organic solvents, 206, 211, 212
- Phenylalanine\*  
Phenylalanylarginine, apparent heats of ionization of, 89  
p*K'* values of, in water, from cells with liquid junction, 85
- Phenylalanylglycine, apparent heats of ionization of, 89  
apparent molal volume and electrostriction of, 159  
dielectric increment of, 146  
p*K'* values of, in water, from cells with liquid junction, 84
- N-Phenylglycine, dielectric increment of, 146
- Phosphate buffers, effect of, on solubility of proteins, 572-3, 597, 604-5
- Phosphoric acid, p*K'* values of, 134
- Phosphorus as a constituent of proteins, 307, 308, 341
- Phycocyan†  
Phycocerythrin†  
p*K* and p*K'* values, *see* Dissociation constants
- Polar molecules, 1, 2, 141  
*see also*: Dielectric constant; Dielectric increment; Dipole moment; Dipolar ionic structure; Relaxation time
- Polarization, dielectric, definition of, 143  
relation of, to dielectric constant, in gases and non-polar liquids, 143  
in polar liquids, 151-2, 294-6  
in protein solutions, 546-7  
in solutions of amino acids and peptides, 149-52  
*see also*: Dielectric constant; Dielectric increment; Dipole moment
- Polarizability, electric, definition of, 142
- Polypeptide chain, structure of, deduced from x-ray studies, 321 ff.
- Potassium chloride, activity coefficients of amino acids in solutions of, 238, 264, 269-70  
effect of, on activity coefficients of proteins, 608 (fig.)
- Potential, of a large thin plane flake between parallel plates, 626 ff.  
of a spherical particle, in a conducting medium, 52-5, 625 ff.  
in an insulating medium, 624
- Precipitin reaction, relation of, to combinations between protein ions, 602
- Proline\*  
Propionamide, partial specific volume of, 376  
solubility of, in water and organic solvents, 206
- Propionic acid, apparent molal heat capacity of, 169  
heats of ionization of, 82, 117  
Raman frequencies of, 11, 15
- Protamines, composition and structure of, 312  
insoluble salts formed with insulin, 575, 601
- Protein-protein interactions, 620-2
- Protein salts, insoluble, 575, 599-602
- Proteins, amino acid composition of, 358-9  
early analytical studies on, 341  
weight per cent of amino acid residues in, 368  
*see* Table of Contents, Chapters 13-25 inclusive
- Proteolytic enzymes, specificity of, and hydrolysis of synthetic substrates by, 309-10
- Pseudoglobulin†, diffusion constant of, 417  
interaction of, with salts, 620  
molecular weight of, from osmotic pressure, 390  
solubility of, 602
- Purity of protein preparations, criteria of, 317
- Pyrazole derivatives, p*K'* values of, 131
- Pyroglutamic acid, heats of solution of, 192

\* *See* Tabular Index of Amino Acids and Peptides, pp. 680-683.† *See* Tabular Index of Proteins, pp. 684-686.

## R

- Rabbit papilloma virus, molecular weight, sedimentation and diffusion constants, partial specific volume, of, 430
- Raman spectra, 10-19  
evidence for dipolar ionic structure from, 12-19  
general relation of, to molecular vibrations, 9-12  
ionization of amino group, studied by, 15-19  
of carboxyl group, studied by, 12-15  
"Regular solutions", definition of, 180  
deviation from laws of, in aqueous solutions, 183
- "Relaxation effect", in electric double layer, 642-3
- Relaxation time, definition of, and relation to molecular size and shape, 506-12  
determination of, from dielectric dispersion, 555-8  
from dispersion of conductance, 558-60  
from double refraction of flow, 532-6  
relation of, to rotary diffusion constants, 508-12  
tables of values of, for amino acids and peptides, 566  
for proteins, 540, 557, 559  
*see also*: Rotary diffusion constant.
- Resonance in molecules, 124-30  
effect of, on acidity of aminobenzoic acids, 127-9  
on acidity of the anilinium ion, 125-7  
on dipole moments, 128  
in the amide group, 129-30
- Ribonuclease†, x-ray data on crystals of, 328
- Rotary diffusion constant, definition of, and relation to molecular size and shape, 506-12  
determination of, from dielectric dispersion measurements, 555-8  
from dispersion of conductance, 558-60  
from double refraction of flow, 532-6  
lengths of protein molecules calculated from, 511, 536-42  
relation of, to relaxation times, 508-12  
significance of, for non-Newtonian flow (anomalous viscosity), 523-7

tables of values of, for protein molecules, 540, 557, 559

*see also*: Relaxation time.

## S

- Salmine, 313; amino acid content of, 358
- Salt formation, between oppositely charged protein ions, 575
- Salting-in coefficient, for interaction between proteins and amino acids, 617-9  
between proteins and ions, 609-17, 619-22  
relation of, to dipole moments, 612-17  
between dipolar ions ( $K_R^*$ ), relation of, to electric moments, 229-35, 617-9  
between ions and amino acids or peptides ( $K_R$  or  $K_R'$ ), 111-5, 245-50, 272-5  
relation of, to electric moment of dipolar ion, 273-5, 280-7  
theoretical calculation of, 277 ff.  
for another ellipsoidal model, 283  
for prolate ellipsoids with charges at foci, 281, 282, 301-3  
for spherical dipolar ions, 280, 281, 299-301, 609-17
- Salting-out coefficient, for interactions between dipolar ions ( $K_S^*$ ), 223, 227, 229-35  
for interactions between ions and dipolar ions ( $K_S$ ), 111-12, 243 ff., 261  
table of values of, 604  
theoretical calculation of, 248, 267-9, 280, 281, 285, 613-4  
*see*: Salting-out
- Salting-out of non-electrolytes, 62
- Salting-out of proteins, 602-9  
effect of pH on, 606-7  
of temperature on, 607-8  
equation describing, 603  
values of salting-out constant, 603-6
- Sarcosine\*
- Sarcosylglycine,  $pK'$  values of, in water, from cells with liquid junction, 85
- Sarcosylsarcosine,  $pK'$  values of, in water, from cells with liquid junction, 85
- Secalin, critical frequency and relaxation time of, 557  
dielectric increment of, 545, 554  
dipole moment of, 554

\* See Tabular Index of Amino Acids and Peptides, pp. 680-683.

† See Tabular Index of Proteins, p. 686.

- Secretin, amino acid composition of: relation to Bergmann-Niemann hypothesis, 316
- Sedimentation constant (*s*), definition of, 422  
 reduction of, to standard conditions (*s*<sub>20</sub>), 422  
 relation of, to molecular asymmetry and hydration, 424-5  
 values of, for proteins, 428-443
- Serine\*
- Serum albumin, horse†, asymmetry and hydration of molecule of, 562  
 dipole moment of, 554  
 interaction of, with salt, 620  
 sedimentation diagram of, 442  
 solubility of, 602  
*see also*: Serum albumin sulfate
- Serum albumin, man, molecular weight of, from osmotic pressure, 392
- Serum albumin, ox, molecular weight of, from osmotic pressure, 392
- Serum albumin, sheep, molecular weight of, from osmotic pressure, 392
- Serum albumin sulfate, horse, solubility and heat of solution of, 587
- Serum euglobulin, horse, density of crystals of, 377
- Serum globulin, horse†, viscosity increment of, 520  
*see also*: Serum pseudoglobulin
- Serum globulin, Lampetra, molecular weight, sedimentation and diffusion constants, partial specific volume, of, 429
- Serum globulin, man†
- Serum globulin, ox, molecular weight of, from osmotic pressure, 392
- Serum globulin, sheep, molecular weight of, from osmotic pressure, 392
- Serum pseudoglobulin, dielectric increment of, 545  
*see also*: Serum globulin†
- Serum  $\gamma$ -pseudoglobulin†, horse, asymmetry and hydration of molecule of, 562  
 critical frequency and relaxation time of, 557  
 dielectric increment of, 551, 553, 554  
 dipole moment of, 554
- Shapes of protein molecules deduced, from dielectric dispersion measurements, 555-7, 560-5  
 from double refraction of flow, 536-42  
 from sedimentation and diffusion, 404-6, 424-5, 434-7  
 from viscosity measurements, 519-22  
 from x-ray diffraction studies, 322-7, 331-7
- Silk fibroin†, composition of, 315  
 frequencies of amino acid residues in, 315  
 x-ray diffraction studies on, 322-3
- Silver iodate, solubility of glycine and alanine in, 253-6
- Smoluchowski's equation, for electrophoretic migration, 628, 640  
 for electroviscous effects, 523
- Sodium caseinate, rotary diffusion constant and approximate length of, 540
- Sodium chloride, effect of, on activity coefficients, of amino acids, 238-51, 256-64, 266-75  
 of peptides, 272-5  
 of proteins, 573, 604, 608
- Sodium sulfate, activity coefficients of cystine in solutions of, 243-6  
 as a precipitant for proteins, 572, 604, 608 (Fig.)  
 salting-out of lactoglobulin in, 604  
 solubility of hemoglobin in, 604, 608 (Fig.)
- Sodium thymonucleate, rotary diffusion constant and approximate length of, 540
- Solid solutions of proteins in crystals, 580, 582-3
- Solubility, effect of particle size on, 51  
 significance of, for protein solubility, 52  
 general relations of, to chemical structure and to intermolecular forces, 177-95  
 of amino acids and peptides, in solutions of electrolytes, 236-75  
 in solutions of other dipolar ions, 217-35  
 in water, at various temperatures, 190-4  
 at 25°, 199-200  
 in water and organic solvents at 25°, 201-4, 206-16  
 of electrolytes in different solvents, 62  
 of proteins, choice of solvents for extraction of proteins, 569-73  
 in water, 574, 586-7  
 influence of ionic strength on, in dilute salt solutions; influence of electric moments, 609-17

\* See Tabular Index of Amino Acids and Peptides, pp. 680-683.

† See Tabular Index of Proteins, pp. 684-686.

- Solubility, of proteins, influence of ionic strength on, in concentrated salt solutions (salting-out), 602-9  
     in solutions of other dipolar ions, 617-9  
     influence of pH on, 575-6, 593-9, 605-7  
     influence of temperature on, 576, 587, 605, 607  
     influence of total mass of saturating body on, 576-87  
     temperature coefficient and heat of solution of, 188-95  
     tabulated values of, for amino acids, 190-2, 204  
     for proteins, 587  
*see also*: Activity coefficients; Heat of solution; Ionic strength; Dielectric constant; Phase rule
- Solubility product of amino acid derivatives, 363-5
- Solubility product constant of proteins in solution, 593-4
- Solvents, choice of, for extraction of proteins, 569-73
- "Specific ion interaction", 64
- Specific volume, *see* Apparent specific volume; Partial specific volume
- Spider silk fibroin, amino acid content of, 358
- Stereochemistry of amino acids, 308
- Stokes' law of viscous resistance to a moving sphere, 402, 406, 624
- Streaming birefringence, *see* Double refraction of flow
- Structures derived from x-ray diffraction studies of alanine molecule, 320  
     of diketopiperazine, 319, 321  
     of fibrous proteins, 322-7  
     of "globular" proteins, 327-32  
     of glycine molecule, 320-1  
     of  $\beta$ -keratin, 325  
     of virus proteins, 332-3  
     of wet and dry insulin (Patterson projections), 336
- Sucrose, apparent molal heat capacity in water of, 169  
     apparent volume and electrostriction of, 163  
     melting point of, 185
- Sulfhydryl group, aliphatic, acidity of, 85, 131  
     pK' values of amino acids and peptides containing, 85
- Sulfhydryl groups, in proteins, presence of, 351, 448
- titratable; increase of, on denaturation, 435, 437, 537
- Sulfur, influence of, in methionine, on relative solubility in water and organic solvents, 206, 211, 212  
     in proteins, 346-52
- Sulfur-containing amino acids, in proteins, 346-52
- Surface concentration of a component, definition of, 50
- Surface energy, 49
- Surface tension, 49  
     of amino acid solutions, relation of, to dipolar ionic structure, 174-6
- Svedberg unit of sedimentation constant, definition of, 422

## T

## Taurine\*

Temperature, influence of, on solubility of proteins, 576, 587, 605, 607

Tetraglycine, dielectric increment of, 146

pK' values of, in water, from cells with liquid junction, 84

relaxation time of, 566

Tetronic acid, pK' values of, 135

Thallous chloride, solubility of, in amino acid solutions, 251-4

Thallous iodate, solubility of, in amino acid solutions, 253-6

## Threonine\*

Thyroglobulin, diiodotyrosine and thyroxine in, 342

Thyroglobulin, pig†

Thyroxine, in proteins, 342

Titration constants (Simms), relation of, to dissociation constants, 449-53

Titration curves of proteins,† 444-505  
     description of, in terms of titration constants, 449-53

effect of ionic strength on, 468-77

methods of determining nature of ionizing groups in, 447-9

temperature coefficients of, *see* Heat of ionization

*see also*: Maximum acid and base binding capacity; Deamination of proteins

Tobacco mosaic virus, double refraction of flow of, 538, 539

rotary diffusion constant of, 540

sedimentation constant of, 430

translational diffusion constant of, 417  
     *see* references under Virus proteins

\* *See* Tabular Index of Amino Acids and Peptides, pp. 680-683.

† *See* Tabular Index of Proteins, pp. 684-686.



- Tobacco seed globulin, density and molecular weight of, from x-ray diffraction studies, 329, 330  
 dimension of unit cell of, from x-ray studies, 329  
 Total charge on ampholyte molecules, distinction from net charge, 93, 446  
 Triglycine\*  
 Triglycine hydantoic acid, apparent molal volume of, 372  
 solubility of, in water and in organic solvents, 201, 206  
 Trypsin, action of, on synthetic peptides, 310  
 molecular weight of, from osmotic pressure measurements, 390  
 specificity of, 309-10  
 Tryptophane\*  
 Tyrosine\*, salting-out constants of, 604  
 Tyrosylarginine, apparent heats of ionization, 89  
 $pK'$  values of, in water, from cells with liquid junction, 85  
 Tyrosyltyrosine,  $pK'$  values of, in water, from cells with liquid junction, 85

## U

- Ultracentrifuge, general principles of design of, 126-7  
 results of ultracentrifugal measurements, 427-43  
 theory of ultracentrifugal measurements, 419-24  
 Unrestricted systems, thermodynamic treatment of, 36  
 Urea, apparent molal heat capacity of, 169  
 apparent molal volume and electrostriction of, 163  
 dielectric increment of, 144  
 dipole moment of, 144, 231, 233  
 influence of, on osmotic pressure of proteins, 389-92  
 on sedimentation and diffusion constants and viscosity of proteins, 436-7, 527  
 on titratable sulphhydryl groups in proteins, 437  
 on viscosity and diffusion of serum albumin, 417, 522  
 on viscosity and double refraction of myosin, 527, 537  
 interaction of, with cystine, 231  
 melting point of, 185  
 partial specific volume of, 376  
 solubility of, in aqueous electrolyte solution; activity coefficient of, 254  
 Urease†

## V

- Valine\*  
 Velocity gradient, definition of, 512  
 effect of, on motion of ellipsoidal particles in a liquid, 517-9  
 on viscosity increments in solutions of asymmetrical molecules, 523-7  
 magnitude of, in flow between concentric cylinders, 513-4  
 in flow through a capillary tube, 513  
 significance of, for double refraction of flow, 532-40  
 Virus proteins, double refraction of flow in solutions of, 537-9  
 liquid crystal formation in solutions of, 332  
 molecular weights of, 428-31  
 x-ray diffraction studies on, 328-9, 332  
 Viscometers and viscometry, 513-4  
 Viscosimeters, *see* Viscometers  
 Viscosity, definition of, 512  
 measurement of, 513-4  
 of solutions, containing large spherical molecules, 514-7  
 of ellipsoidal molecules, 517-27  
 of solvent, influence of, on electrophoretic migration velocity, 624-43  
 on rotary Brownian movement, 510-2  
 on sedimentation velocity, 422-5  
 on translational diffusion and Brownian movement, 402-6  
 relation of, to molecular asymmetry and hydration, 519-22  
 values of, in myosin, 525-7  
 in tobacco mosaic virus, 524-5  
 variation of, with change of velocity gradient, 523-7  
 Viscosity increment ( $\nu$ ), definition of, 515  
 of spherical molecules, 515-7  
 relation of, to molecular shape and hydration, 519-27

## W

- Wool protein, combination of, with acids and bases and with acid anions, 495-8  
*see also*: Keratin

## X, Y, Z

- X-ray diffraction studies of proteins, 5, 318-37  
*see also*: Structure derived from x-ray  
 Yellow enzyme†  
 Zein†  
 Zinc chloride, effect of, on activity coefficients, of amino acids, 264-5  
 of proteins, 620

\* See Tabular Index of Amino Acids and Peptides, pp. 680-686.

† See Tabular Index of Proteins, pp. 684-686.





lysine.....	79-80	84, 99, 109, 121-2	107, 109	137	79, 80, 89	144, 146 148-9, 545	121, 141, 231, 238, 275, 287, 290, 547	412	566	185, 188	15	315, 358-9, 368
lycylglycine.....		84, 99, 113-4			89	144, 146, 150, 290, 545	144, 231, 233, 275, 290		566	185		
histidine.....		85, 87		137	89							315, 354, 358 368, 499
hydroxyproline.....		84		137-8		148						358-9
isoleucine.....	79-80	84			79-80			412				358-9, 368
eucine.....	79-80	84, 99	107	137	79-80	146	287	412		185		358-9, 368
lysine.....		85		137	89							315, 354-5, 358-9, 368, 499
γ-glutamic acid.....		86			89	146, 545	231, 233, 275	412	566			348-50, 358-9, 368
β-thionine.....		84										
phenylalanine.....		84		137		146, 148		412		185	15	358-9, 368
proline.....		84		137		146						358-9, 368
arabosine.....		85		137-8								358-9, 368
serine.....		84										
taurine.....		84				146						
threonine.....		84										358-9, 368
triglycine.....		84				146, 150, 545	231, 275		566			
tryptophane.....		84		137-8				412		185		358-9, 368
tyrosine.....		85		137								315, 358-9, 368
valine.....	79-80	84		137-8	79-80, 89	146	233	412		185		358-9, 368

# Tabular Index of Data for Certain Important Amino Acids and Repurposes. Part II.

(Further references concerning many of these substances are given in Part I and in the Subject Index)

	Solubility and Activity Coefficients in							Heats of solution	Apparent molal heat capacity	Apparent molal volume and electrostriction	Surface tension increment
	Aqueous electrolyte solution	Electrolyte in ethanol-water mixtures	Ethanol-water mixtures	Organic solvents	Solutions of other dipolar ions	Water at 25°C	Water at different temperatures				
-Alanine.....	252, 254, 256-7, 259-60, 264		111, 203, 215	199, 201, 203, 206, 208-9, 201, 203, 206	220, 618	198-9, 203, 206, 208-9, 219, 221	190	192, 194	169, 171-2	158-9, 163-4, 372	174-5
-Alanine.....	254, 257		203			198, 203, 206, 228			171, 172	158-9, 163-4	174-5
minobenzoic acids (o, m, p).....											
-Amino-n-butyric acid.....	252, 254, 257, 259-60, 264		203, 215	199, 203, 206, 208-9	220, 618	198-9, 203, 206, 208-9, 219, 221		194		159, 163, 372	174-5
-Amino-iso-butyric acid.....	254, 257		203			198-9, 203, 219, 221		194		159	
-Aminobutyric acid.....											
-Amino-n-caproic acid (Norleucine).....											
-Aminocaproic acid.....											
-Amino-n-valeric acid.....	254, 257		111	206		198, 228				159	174-5
-Arginine.....	241, 262, 266		111, 203	199, 203	230, 223-4, 229, 231, 233, 241, 610	199, 203, 206, 208-9, 273	190	192		159, 372	174-5
-Sparagine.....						198, 203, 228				159, 163-4	

[illegible]

## Tabular Index of Data for Certain Important Proteins. Part I

(Further references concerning many of these proteins are given in Part II and in the Subject Index)

Proteins	Amino acid content			Acid and base binding capacity	Titration curve	Dielectric increment	Relaxation time	Viscosity increment
	Sulfur containing amino acids	Basic amino acids	Other amino acids					
Casein.....	348, 358	354, 358	358-9	354, 459				
Coconut globulin.....	358	358	358					
Edestin.....	348-9, 358, 368	354-5, 358-9, 368	358-9, 368	354, 459		545, 554	557	
Egg albumin (ovalbumin)	348-9, 358, 368	354, 358-9, 368, 499	358-9, 368, 499	354, 455, 498-500	469, 499	545, 552, 554, 559	557	520, 522
Fibrin.....	348, 350							
Fibrin, cattle..	358	355, 358	358-9					
Gelatin.....	348-9, 358, 368	355, 358, 368	358-9, 368	355, 459	493, 495			
Gliadin.....	348-9, 358	354, 358	358-9	354		545, 554	557	520
Hemocyanin*..	348	355						520
Hemoglobin, horse.....	348, 350, 358	355, 358	358	355, 478	466, 480, 484	545, 550, 554, 559	557	520
Insulin.....	348-9, 358, 368	354, 358-9, 368	358-9, 368	354	502	545, 554	557	
Keratin, wool..	358	358	358					
$\beta$ -Lactoglobulin.....		354-5, 501	501	354, 500-1	470, 500	545, 554, 559	557, 563	520
Myoglobin....						554	557	
Myosin, rabbit.	348, 350	355		355				525, 527
Pseudoglobulin- $\gamma$ .....		355		355				
Serum albumin, horse.....	348, 350, 358	355, 358	358-9	355, 504	503-4	545, 554	557	520, 522
Silk Fibroin....	358	315, 358	315, 358					
Thyroglobulin, pig.....	348							520, 522
Zein.....	348-9, 358, 368	354, 358, 368	358-9, 368	354, 490-1	490-1	545, 554	557	

\* Various species.

**Tabular Index of Data for Certain Important Proteins. Part II**  
(Further references concerning many of these proteins are given in Part I and in the Subject Index)

Proteins	Partial specific volume	Dimension of unit cell from x-ray data	Density of crystals	Molecular Weight from			Translational diffusion constant	Sedimentation constant
				Osmotic pressure	Sedimentation and diffusion	Sedimentation equilibrium	X-ray diffraction	
Amandin.....	429			390-1	429	429		416, 429, 520
Antipneumococcus serum globulins*...	428-30				428-30			428-430
Bence Jones- $\alpha$ .....	428					428		428
Bence Jones- $\beta$ .....	428				428			416, 428
Bushy stunt virus.....	430	329	329-30		430	430	329-330	417, 430
Catalase.....	429				429			416, 429
Concanavalin A.....	428				428			416, 428
Concanavalin B.....	428				428			416, 428
Crotoxin.....	428				428	428		428
Cytochrome-c.....	428				428			416, 428
Diphtheria antitoxic pseudoglobulin.....	428-9				428-429			428-429
Edestin.....	375, 379, 429		377, 379	391	429			416, 429
Egg albumin (ovalbumin).....	375, 428		377	385, 390-1	428	428		415-6, 428, 520, 522
Erythrocytorins*.....	428, 430				428, 430	428, 430		416, 428, 430
Excelsin.....	429	329	329-30	390	429		329-330	416, 429
Gelatin.....	375							
Gliadin.....	428			390-392	428	428		416, 428, 520
Hemocyanins*.....	429-30			390-391	429-430, 541	429-430		416, 429-430, 520
Hemoglobin, horse.....	379, 428	329	378-9	385-386, 389-392	428	428		415-416, 428, 520

\* Various species.

## Tabular Index of Data for Certain Important Proteins. Part II (Concluded)

Proteins	Partial specific volume	Dimension of unit cell from x-ray data	Density of crystals	Molecular Weight from				Translational diffusion constant	Sedimentation constant
				Osmotic pressure	Sedimentation and diffusion	Sedimentation equilibrium	X-ray diffraction		
Hemoglobin, man.....	428			392	428			415-416, 428	428
Hordein.....	428				428			428	428
Human tuberculosis bacillus protein...	428				428			428	428
Insulin.....	375, 428	328, 336	328		428	428	328	416, 428	428
Keratin, wool		325							
Lactalbumin..	428				428			416, 428	428
$\beta$ -Lactoglobulin.....	428	328	328-9, 380-1		428	428	328-329	416, 428, 520	428
Myogen .....	429				429	429		429	429
Myoglobin.....	428				428	428		416, 428	428
Nucleohistone.	430				430	430		430	430
Pepsin.....	428	328	328	390	428	428	328-329	416-7, 428, 520	428
Phycocyan ....	429				429	429		416, 429	429
Phycocerythrin.	429				429	429		413, 416, 429	429
Ribonuclease..	428	328	328		428	428	328	417, 428	428
Serum albumin, horse...	428	329	329, 377	386-392	330, 428	428	329-330	415, 416-7, 428, 520, 522	428
Serum globulin, horse.....	428			390-392	428	428		416, 428, 520	428
Serum globulin, man.....	428			392	428			428	428
Thyroglobulin, pig.....	429				429	429		416, 429, 520-2	429
Urease.....	429				429			416, 429	429
Yellow enzyme.	428				428	428		416, 428	428
Zein.....	375, 428			390-391	428			428	428

\* Various species.

